ACIT5930

MASTER'S THESIS

in

Applied Computer and Information Technology (ACIT)

May 2023

Cloud-based services and operations

Approaching online gaming patterns through correlations and similarity

May Christin Fevang Johansen

Department of Computer Science
Faculty of Technology, Art and Design

OSLOMET

Acknowledgements

I want to express my deepest gratitude to my supervisor, Kyrre Begnum, for the support and guidance throughout this project. The knowledge, ideas and advice he has provided throughout the process of this project is invaluable to me. I appreciate that he has been available through this entire project when needed, as well as being extremely helpful. The guidance process has kept me motivated and excited for the work in this project. Thank you so much.

I would also like to express my gratitude towards my family and friends, for their support and believe in me when it comes to working on finishing this project. It has been time-consuming, and I appreciate the continuous understanding that has been shown to me. Lastly, I would like to thank my husband for always being there for me, by showing his support every day and discussing video games with me when needed.

Abstract

The purpose of this master thesis is to investigate similarities found in online video game patterns, with the use of applications found in Steam, a distributor of video games. This has been done through correlation in order to investigate what video games are found to be similar, or dissimilar. The research has been done by creating a few assumptions and hypotheses, which serves as a foundation for the results and findings. The video game industry is a billion dollar industry and there are a lot of factors surrounding it that makes it important to conduct further research, in order to discover potential for new development. The approach has been to investigate the video games both in a narrow and broad approach, which combined has allowed for more knowledge on the subject. In the cases were similarity were presumed, it has been found, but it is not a bulletproof method. There is still need for further research on this subject, but the findings in this project can contribute to future work.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Today's society is undoubtedly influenced by technology and the aspects related to the field. All the different changes in the way we live, work and relate to one another is considered the fourth industrial revolution. This era is enabled by extraordinary advances in technology, and it has also contributed to changing the way we think, act and learn [1]. The business industry today, is immensely affected by the advances and development that are made within the technological field. With the huge impact that these advances can have, it is crucial to stay updated on different trends and try to optimize the opportunities that a company can have moving forward in time.

There are a lot of different industries in the world, for instance agriculture, aerospace, transport, telecommunication, computer, gaming industry and more. All of these are being more and more digitalised and are strongly dependent on technology. During the course of the last 20 years or maybe even more, the society in general has been more digitalised and a lot of different things are heavily dependent on technology and enormous systems. A lot of our internet usage is very large and complex systems.

Complex systems can be found everywhere, in everything from companies that work in web development, to a sports team or a company that develops video games. For instance, a supermarket is just as dependent on technology as a hospital or a doctor's office. When a supermarket loses its power it is quite severe because everything is dependent on technology, the register, bank terminals, and refrigerators and freezers. So when something happens the only solution may be to work on getting the power back on, in addition to this the bank terminals are also dependent on internet connection, and without it makes the process for paying much more complex and time-consuming. This is just a minor example, which is important but not life threatening, and to some extent it shows that the smaller aspects of life that we might take for granted also is a big system that needs to be up and running most of the time.

Another example is the Facebook outage on October 4th 2021, where Facebook and other social medias like WhatsApp and Snapchat had problems for multiple hours.

This event had an enormous coverage in the media all over the world. The outage did not affect all countries in the same extent or severity, and the reason for this is that some countries and governments use Facebook more than others. In some countries, especially developing nations, there is a high dependency on Facebook and WhatsApp when it comes to the free messaging services. This applies to small businesses and informal economies that heavy rely on the services that Facebook provides. In addition to this there are tens of millions of businesses that use Facebook, Facebook Messenger and WhatsApp for their e-commerce sales. These messaging services are being used for communication between sellers and buyers. In Europe and North-America there are other options for communication, but this is not the case for developing nations like India, Mexico and Brazil [2].

Because of this, it is important to have the ability to understand how these types of systems work and the impact it has on a business, government or a community. It might not be as recognizable to everyone why this particular event is important to discuss and remember moving forward in time, especially for citizens of Europe and North-America as previously mentioned. With being aware of a situation like this particular outage, it is possible to be better prepared or maybe even be able to prevent that something similar does not happen again. Of course, that is not to say that Facebook or any other social media will not have an outage again, but with identifying patterns one can use the data to ones advantage in the future.

It is important that large and complex systems are stable, and this is something that might be taken for granted in many situations. For instance, if there is downtime on a webpage or an app one might get frustrated and maybe also annoyed because it does not work. The question is whether this is reasonable or not? Should we expect that something works one hundred percent of the time? There are many factors that needs to be taken into consideration when trying to answer or discuss these questions. In an ideal world we would want to make sure that our systems are constantly stable. It might be easy to think that we can just use endless money and resources on keeping things in the desired state at all times, but it is not that straightforward. The reason for this is first and foremost that it is not a sustainable way of working, and in many situations we cannot be on standby or have backup solutions for everything. It might be a good idea to take extra precautions around

certain events. One example of this can be for a company that does web development and especially in the e-commerce industry around Black Friday. This is an event that will generate a lot of traffic and purchases in an online store, and it can be considered necessary to be prepared for something to go wrong. The reason for this is that a potential outage will do a lot more harm during a special event like this.

The video game industry is also an example of an industry where the systems preferably needs to be stable at all times, but as with other systems or industries it is not guaranteed to always be one hundred percent stable. As with the previous example of Black Friday, we can take into comparison when there is a new video game or new downloadable content (DLC) for a game being released. With popular games there is often a considerable amount of anticipation leading up to the release of a new game or a DLC. In these scenarios it can also be in everyone's best interest to be prepared. This particular industry is an example of something that has evolved significantly as technology continues to improve. Creation of video games has become more and more complex, and the cost related to creating a video game can cost tens and maybe even hundreds of millions at this point in time. This sector is continuously growing and there are more than two billion gamers around the world [3].

Releases of new games and events like Black Friday are dates that are already known to us, and therefore it is easier to prepare for them again and again. This way, systems will be available but at the same time adaptive. So how do we deal with other events and aspects that are unknown to us? We can do this by trying to determine patterns, and when these patterns are established we can monitor them in order to have more predictability moving forward. These patterns can be used both to establish whether or not we need to look for something specific, or if we want determine if something acts or looks the same.

In regards to the video game industry, the purpose is to analyse data relating to two thousand applications that are available through Steam, which is a game distribution platform. The dataset consists of a quantity of information, for instance, price and player count history. By analysing this kind of data, and maybe visualizing it in graphs and other formats, it can be used by game developers and others in the future, for

example when developing a new video game. When information is visualized, it is easier to identify patterns and trends rather than looking through multiple documents or spreadsheets [4]. If we then combine a particular set of parameters and visualize it, we can conduct an analysis and use our findings for a specific purpose.

In order to analyse data it might be beneficial to look at how big data and business intelligence can be combined. Business intelligence is a term that is already associated with analysis and the presentation of business information, simultaneously the term is evolving and improving, driven by big data, cloud computing and advanced analytics. The combination of technologies and methods have opened up for the possibility of engaging in new use cases that could not even have been considered earlier [5]. In pursuance of achieving the desired solution and result, we need to look at the data input that is provided and other data sources that is available. The question here is if we want to find more information to make our sources even more complete or if the desired analysis is possible to conduct either way.

Similarity in video games is an important subject as it can contribute to learning more about the industry, and about the people that play the video games. This is beneficial for the video game developers, but also for an individual that plays these video games, as it can contribute to a better overview of the industry and choices that exists, which again lead to easier navigation within the industry. Research have been conducted on similarity within avatar making in video games, and Trepte and Reinecke (2010) states that “it is plausible to assume that players are more likely to identify with a game character, if they generally support the game’s narration or wish to be like the character” [6]. This is an interesting aspect of the industry as avatars in video games can be very important, especially when playing a MMORPG.

There are several ways to investigate if games are similar or have some extent of similarity, but the question is in regards to what is the best approach with the data that is available. The analysis can be done on genre, price history, amount of players or maybe games that are free to play. The entire aspect is dependent on what the desired goal actually is, and what the results of the analysis can or should be use for

at a later stage. Conducting analyses and using the data that we already have access to is important. This is crucial in a society that is continuously in development and especially when it comes to technology and digitalization, companies in all kinds of industries can have great success if they stay updated on the latest developments within technology and possibly also trends within artificial intelligence.

Artificial intelligence (AI) is an aspect that becomes more and more important in video games. Due to the fact that AI can contribute to enhancing the experience for a player, while at the same time AI can also help the developer in a way that they do not have to build every element of a game. When you are in game and are faced against a non-playable character (NPC), AI is used to create behaviour that is responsive, intelligent or adaptive. For instance, if you are in an in-game fight towards a NPC, and you hide from them for too long the NPC will be able to use other methods to try to find you. This way, AI is an important part of video games as it enhances the experience for the player, rather than machine learning or decision making can do.

There is a considerable amount of video games available today, and it is something that has become increasingly more popular throughout the years. People from all different age groups can be found in the gaming community, as there are so many different video games available and something that is suitable for everyone. This also means that people can be found playing brand new video games, video games that are must established and maybe even revisiting old video games, also remastered editions. We also know that games can be purchased or accessed through multiple platforms or distributors, and Steam is one of them. Steam offers a broad selection of games in different genres, so how can we determine whether or not some of these are similar? Similarity through correlation and principal component analysis (PCA) is a logical starting point. PCA is a statistical procedure which allow you to summarize information in large data tables by means of a smaller set of “summary indices” that can be more easily visualized and analysed [7].

1.1 Problem statement

- *Investigate similarities found in video game usage*

There are several different things that can be done to investigate these similarities, and if something is similar or not can also depend on different aspects. This means aspects related to the video games themselves, what clustering algorithm that will be used and the statistical methods that are applied.

Chapter 2 Background

This chapter will contain information regarding the purpose of the project, as well as technical concepts that will be of importance. Moving forward, it will be important to mention different aspects of the video game industry as it is growing in a rapid pace.

2.1 Introduction to video games and the industry

A video game is an electronic game that can be played on a computing device, for instance a computer, gaming console or a mobile device like a phone or a tablet. The first video games were developed and released in the 70s, and have evolved immensely in the last decades [8]. In 2021, this industry was predicted to be worth 178,73 billion dollars, which is an increase of 14.4 percent since 2020. In 2016 the total worth of this industry were forecasted to be 90.07 billion dollars, at the difference between the forecast and actual worth is 76.8% difference. At this point in time, the worth is forecasted to be worth 268 billion dollars by 2025 [9]. These forecasts show how much and how fast the video game industry is evolving.

There is a broad range of video game development companies and distributors, everything from small companies that just develops one game, to the bigger companies that are well known, a few examples are Blizzard Entertainment, Valve Corporation, Epic Games, Electronic Arts and Nintendo. Blizzard Entertainment, Electronic Arts and Nintendo usually provide games that they have developed themselves, through respectively Blizzard Battle.net, Origin and Nintendo Store. On the other hand, Valve Corporation and Epic Games, have their own distribution platform. Valve is the developer of Steam, and Epic Games have their platform with the same name as the company. Valve has also developed some of the most played games that are accessible through Steam, amongst them are *Counter-Strike: Global Offensive* and *Dota 2*.

Steam was released by Valve in 2003 and as already mentioned, it is a digital distributor of video games. This client works in a way that people can purchase

games and install them to their cloud drive. Steam also consists of reviews of the games, and you can both chat with your Steam friends and create a Steam community which is similar to a discussion forum. Steam is available on Windows, iOS, Linux, TV and mobile devices [10].

2.2 The video game community and aspects surrounding it

As previously mentioned video games has been around for several decades and not only is it a billion dollar industry, but it is also a contributor to great communities and a social highlight for many people. The first example of this is that in many massively multiplayer's (MMO) there is a low threshold for asking for help as there often is a general chat function where you can ask questions and the people that are in the same area as you are able to see those questions. From there you easily get in touch with someone that is a complete stranger and in most experiences people are kind to one another and try to help to their best ability. There are also video games that provide an in game voice-chat, especially those games where you are playing for the same team like Counter-Strike, Overwatch or Call of Duty. Those games and games that are similar to them have quite a high-pace and it is therefore favourable to be able to talk rather than having to type out your message.

There are also some MMO's that have the ability to create or join a guild. The definition of a guild is that it is "an association of people with similar interests or pursuits, especially a medieval association of merchants or craftsmen" [11]. Guilds can be found in video games like World of Warcraft, Guild Wars and New World. The purpose of a guild can be several different things, most common to be able to go through dungeons or raids together. There are also social reasons for creating or joining a guild, as this can be to both create a guild with your friends or to just find people with the same interest or background as you. Background in this case could mean a guild that is focused on having members that speak the same language, maybe a team within e-sports, the possibilities here are many. Besides that, some might join a guild for social reasons but end up with new friends or in some cases there are people who have found love through video games.

This shows that these types of communities is extremely diverse and that means that no matter who you are or what you do, there is a place for you somewhere. The unfortunate side of this is that the gaming community can in some ways be looked down on, mostly because it is misunderstood in many ways. You often see articles or debate post from people in different newspapers with the concern that video games are not only violent and contribute to violent behaviour, but that it takes away time and other opportunities from someone's social life. For instance, in a debate post from 2019 Ole Andre Bråten says that his advice is to turn off Fortnite, as young people might be using a platform where they learn aggression [12]. Johannes, Vuorre and Przybylski (2021) published at paper with the title "Video game play is positively correlated with well-being, where they state that they "deliver much-needed evidence to policymakers on the link between play and mental health [13], which is on the contrasting side of the violence and aggression discussions. As already mentioned video games is quite social and might be even more social for some people than what playing outside is. The reason for this is that some people might be left out or maybe even bullied at school, but them you see that they find their place and are much more comfortable inside a virtual world.

2.2.1 Specific example of the video game community: Komplett

Another great example of the impact and unity that exists within the video game community is a stream initiated by Komplett that for the first time took place in November 2020. Komplett is a Norwegian e-commerce company that was established in 1991 with their headquarters in Sandefjord. They did a charity stream on Twitch in conjunction with Black Friday called "Gamere mot Barnekreft", where the funds they raised where donated to Barnekreftforeningen, which is the Norwegian paediatric cancer association. The stream itself lasted for 27 hours and they collected two and a half million NOK, where the first million were collected in only a few hours and that only consisted of donations from normal people watching. After the success of the first stream in 2020, Komplett has repeated the event and hosted a stream for the same cause both in November 2021 and 2022. In total, they have raised six and a half million NOK for this cause. This exact event is a prime example of how much engagement and care you can actually find in such a great community as this, and show their kindness and support for important causes. We already know that the

video game industry is big and probably bigger than most people think, and events like this might also help to remove the stigma towards gaming that is found in some places.

Barnekreftforeningen published an article on their website on December 3rd 2020 which was a summary of the first event. They could never have imagined that the stream would generate so much money, and therefore the joy and engagement were even bigger. They also emphasise that gaming is no longer a sport for just one person where you are isolated in your own world. Instead it has become a way for children and teenagers to keep in touch with their friends and the outside world in general. For children that have cancer or other illnesses, gaming is a huge contributor to helping them being social when they have to be isolated in a hospital for instance. Gaming and streaming communities also become kind of a safe haven or a sanctuary, despite the seriousness of their life-situation. In addition to this, because physical games or activity can be challenging, video games still contribute to a sense of achievement and unity because they don't set the same physical limitations [14].

Komplett is a mainly know for the fact that they are well established within the e-commerce industry, and their main focus are gamers and video games even though they have a broad selection of other electronics in their store. They often have streams on Twitch in a smaller capacity than mentioned earlier, for socializing and playing video games. In addition to this they have a lot of different campaigns that targets gamers, both with information about new technology and special offers for different products. They are also a sponsor of a Norwegian community called "Nerdelandslaget".

2.2.2 Specific example of the video game community: Nerdelandslaget

Nerdelandslaget is first and foremost a podcast started by Stian Blipp and Andreas Hedemann in April 2019, but it has become a big community especially for gamers in Norway. Their focus is mainly their love of video games, but also to normalize video gaming and provide a community for likeminded people. They release one episode each week alternating between featuring a guest, usually a celebrity that has some

sort of background with gaming, or someone who works within a field related to video games and episodes where only the hosts of the podcast are present. The podcast has become more and more popular, and they have added more hosts which are Sebastian Brynestad, Ida Horpestad and Hasse Hope. In addition to the podcast they have an account on Twitch, where they occasionally stream different games and host different events. Nerdelandslaget have also created a Discord server which goes by the same name as the podcast, and it has more than 8500 members [15].

There are several different purposes of the Discord server, amongst them are socializing and gaming. Additionally, they publish information about the upcoming guests in the podcast episode and different tips from the podcast episodes, as films, video games, tv shows and books often are discussed. As mentioned there is also a social aspect to this community, and they have aimed to create a community that is a safe haven for everyone. As discussed earlier, video games is something that can be a really inclusive tool between people, both strangers and friends. The discord server by Nerdelandslaget have different text and voice channels, where everyone can join and ask questions, get help and tips, or maybe even some new friends that play the same game. It is no doubt that is a prime example of the significance video games have for some people, and it is unquestionably important for individuals to feel a sense of belonging.

The social aspect of video games comes in various forms, and the diversity it brings is what makes it so essential. It can be everything from being engaged in video games and having discussions about them, to playing on a Nintendo Switch or other handhold console, to meeting up and playing multiplayer or even take turns with a controller or keyboard and mouse if it is a pc game. There are so many different ways of approaching gaming, which again speaks to the fact that there is something that is suitable for each and everyone.

2.3 Cloud gaming

The video game industry has been around for many decades already, and since it is a technological invention, it is therefore crucial to follow along with breakthroughs and advances in technology. One of these advances is artificial intelligence, and is

something that is familiar to most people. The availability of music, tv shows and movies today, is something we associate with as on-demand and streaming. Streaming is something that is possible for the reason that cloud computing exist, and some video games are also adopting this concept. Due to the fact that we have cloud computing, it allows video games to run on remote servers and at the same time it is possible to steam directly on their own device [16].

As of today, the concept of cloud gaming exists in a very small scale. This is a huge potential within cloud gaming, but as a result of delays and other concerns, if has not been a success for serious and professional gamers. The benefits with having video games on demand is for instance that it does not require the most modern and advanced hardware for one to experience the best video games. More and more of the leading technology companies, such as Google with Google Stadia, Sony with PlayStation Now, Microsoft with Xbox Game Pass and NVIDIA with GeForce Now have joined the cloud gaming market with new products. In addition to this, Microsoft, Amazon and Electronic Arts are introducing their own cloud gaming platforms, respectively Project xCloud, Amazon Luna and Project Atlas [17] [18] [19]. Domenico et al. (2021) have recently done network analysis on Stadia, GeForce Now and PlayStation Plus, which states that Stadia and GeForce Now are able to stream data up to 44 Mbit/s, and they compare this to being much higher than a 4K Netflix film. They also suggest that there is a lot more work to be done when it comes to these services, and that xCloud and Luna will have to be a part of further analysis [20]. This speaks to the fact that cloud gaming still is brand-new, and demonstrates why cloud gaming is not acknowledged or applied on a broad spectre at the present time.

Furthermore, there are some papers about video games and disorders, for instance Compañ-Rosique et al. (2019) with their paper on making video games accessible to users with cerebral palsy [21]. There is also a survey by Jaliaawala and Khan (2019), which questions if autism can be catered with artificial intelligence-assisted intervention technology [22]. These two papers are in many ways quite contrasting, but the message is essentially how broad the video game industry and the importance of the many possibilities that comes with it. Undeniably, these are quite broad issues which definitely does not come with a straightforward solution or path.

Nevertheless, it is large numbers of aspects that can be used moving forward in time. It may be easier to take into consideration the possible outcome, when one is already more aware of how broad this industry actually is.

2.4 Big Data from a Human-Computer Interaction perspective

Big data and human-computer interaction (HCI) are two concepts familiar to everyone that is working in or studying a field related to information technology. Big data is a combination of structured, semi-structured and unstructured data, that is gathered by organizations in order to do research on. This type of data can be used for several analytical methods or in machine learning projects. The characteristics of big data is the 3 V's, in which refer to the large volume of data, the wide variety of data and the velocity of data [23]. Different industries can use the big data that is acquired in their system for many purposes, for example improve operations and create personalized marketing campaigns based on the data that already exists in their system. This is something that is in use by some video game distributors, as well as Netflix and other streaming services, where you often get new suggestions based on your previous actions.

HCI is a multidisciplinary field of study which focuses on how people interact with computers and whether or not the computers are developed in order to interact with humans in a rewarding manner. The name of the field consists of three parts, which is the user, the computer, and lastly, the interaction between them. It is obvious to us that there are several differences between humans and computers, even so, HCI has an important role in order to make sure that the two parts can interact with each other. When taken into consideration what is known about humans and what is known about computers, it can be easier to develop a system for the specific user that it is intended for.

As HCI is about humans and computer, it is also relatable for the concept of video gaming. This is because video games revolves around a human using a computer, whether that is a PC, console or other handheld device. As Barr, Noble and Biddle states (2007) video games "are software, running on computers, used by people via an interface." Still, the focus within video games are not the same as traditional HCI,

which is about tasks defined by the user, because video games dictate tasks for the players [24]. More research HCI and video games combine can contribute to better understanding and marketing surrounding video games and services.

Further, there is a new discipline called big data analytics that is forthcoming. This discipline is based on a workflow that distils terabytes from low-value data, into a single big of high-value data. The intention with this is to see the big picture based on the information that exists. Big data is advantageous when it comes to research for both HCI and user interface design [25]. For instance, when it comes to A/B testing, which is a method where two versions of a web page or app is compared in order to detect which one performs better. This kind of testing is basically about showing two or more alternatives to random users, and then perform statistical analysis to detect which one is better for the given goal. These kinds of tests were previously evaluated in strict laboratory conditions, but it is now possible to implement these tests quicker and on a larger selection by running controlled experiments [26]. This is something that possibly could be adapted for the video game industry as well, so that we can use the data that we already have for multiple purposes in the future. In addition, it shows that it is important to take measures in order to conduct analysis in a more efficient way.

According to Elmqvist and Irani (2013), the rise of big data increasingly demands that one has access to data resources at any given time or place in order to support decisions and activities such as travel, telecommuting and distributed teamwork. Furthermore, there is an approach that is referred to as ubiquitous analytics, or ubilytics. This approach is based on using multiple networked devices in our local environment to facilitate broad and dynamic analysis of massive, heterogeneous and multiscale data. Ubilytics can be beneficial because the analysts are not bound to their office, and it can contribute to deep cognitive processes which can advance the process of analysis [27].

2.5 Human-data interaction

As a result of the growing trends within big data, new data collection and other interaction techniques, a new field is being established. This field is called human-

data interaction (HDI), and is likely to lead to new opportunities in the future, both for analysts and for designers. With HDI, the outline is that the human should be at the centre of the flows of data, and from there be able to provide components that allows citizens to interact with these systems and this particular data explicitly.

Related to this field, it might be important to study the distinction of whether data is created by us or if it is created about us by others. In general, the interaction with individual computing systems, such as social media is relatively conscious, but when it comes to other computing systems where there are multiple parties involved, one might not always be conscious of this [28]. According to Mortier et al. (2015), “increasing practice of accumulation of data and the increasing importance of these data and inferences drawn from them for our every lives drives the need for the study of HDI” [29]. This is something that is important when it comes to payment solution providers, in which there are several available through all the different video gaming distributors. There is most likely an increasing number in the amount of video games that are being purchased through a distributor rather than the video games being purchased in a physical store. These kind of payment providers often have built-in logic to store information like a card number, so that the only action that needs to be completed is to provide the card security code (CSC) for each check out. This can contribute to easier accessible purchases and more sale for the video game companies or distributors. This definitely constitutes a user friendly solution, but can cause disadvantages when the data is stored and the completion of purchases are more available.

2.6 Adaptive User Interaction

User interface (UI) is a familiar subject within information technology, and it revolves around the HCI and communication in a device. This can for instance include screen, keyboard, mouse or the desktop, through these devices the user interact with an application, website or video game. Because of the growing availability and need to be visible online, the importance of providing the user with a good experience has become more and more important [30]. In addition to the normal user interface, there are more complex ways of creating an UI. For example, adaptive user interfaces is a type of interface that can change their appearance or interaction behaviour, in order

to match the user, device and context. An adaptive user interface should be able to perform changes automatically according to the available information about the user, device and context [31].

The user interfaces of interactive systems that are available today are becoming more and more complex, and therefore it is no longer acceptable to provide a user interface that is suitable for everyone. Building multiple user interfaces with the same functions but suited for different contexts will also cause problems because content often changes, and it will be too time-consuming and higher costs will occur. Therefore, solutions that can be beneficial needs to be researched further. Yigitbas et al. (2019) refers to model-driven user interface development (MDUID) approaches that previously has been proposed in order to make the development of user interfaces more effective. There are some challenges related to MDUID in order to support the development of a self-adaptive user interface, but it is still a promising candidate in order to comprehend the complex task of developing a self-adaptive user interface [32].

An adaptive user interface could be beneficial when it comes to developing a tool for analysis, and can also be used in this project as the results and patterns from video games and player counts might benefit from being displayed more visually, especially when it comes to moving forward in time and finding a purpose of the results. There are multiple data sources which can be taken into consideration and they can serve a different purpose moving in the future. It might also be a different intention or purpose of the analysis and this demands an user interface that can administer the complexity this leads to.

2.7 Adaptive or adaptable user interface

Adaptive user interfaces (AUI) are able to facilitate the handling of computer systems through the automatic adaption to users' needs and preferences. This can also have a greater impact on this user experience itself, as it might lead to greater satisfaction for the user. Adaptive user interfaces is also referred to as Intelligent User Interfaces (IUI) in technical literature. This type of interface can be considered a multidisciplinary field and it is based on research from different disciplines, the most

important of them being artificial intelligence, user modelling and human-computer interaction. Other disciplines related to IUI's are psychology, ergonomics, human factors, cognitive sciences among others [33]. Within AI, adaption and problem solving are some of the basic elements when it comes to research. Therefore, a lot of IUI's are adjusted to use AI techniques, but not all of them have the capability to learn or solve problems. IUI's are designed to improve interaction and communication between the machine and the user. Looking at this from a HCI perspective, it is not prominent what techniques are used to achieve improvement but the actual improvement itself that is the important factor.

It is not a new concept that a system is able to adapt itself based on requirements or requests from the user, and literature shows that there are many approaches in order to design flexible user interfaces. Further, the design of flexible user interfaces can be divided into two categories: adaptive and adaptable [34]. The difference between these two is that adaptive systems adjust their display and available operations by monitoring user status, the state of the system and particular situation. On the other side, adaptable systems is based on providing alternatives for the user. The difference between these two are miniscule, and both approaches have their advantages and disadvantages. These advantages and disadvantages needs to be measured towards the actual goal and purpose of the user interface one might desire to develop.

One of the questions at hand is whether or not these technologies and concepts can be adapted by a more analytic side of the video game industry. It is already known to us that video games can to some extent be compared or classified just by knowing the genre of the video game, as there are a lot of different genres and subgenres. Regardless, what we want to research further is if there is other aspects that makes video games similar. In addition, we might want to know what those aspects are so that we can use them as an advantage moving forward. Data can possibly be displayed in a user interface, the reason for this is that, for many it is easier to make use of the data when it is presented visually as that might seem more inviting as well.

When a specific goal is set it is easier to find the right path moving forward, as it is necessary to see all the possibilities on context with that particular goal. When we

have more information about the specific task at hand, or at least the general approach it will be more clear as to how the data and information we have can be combined with other specific technologies in our research. This also relates to what kind of user interface that will be most beneficial for the industry. The video game industry consists of a multitude of data and data sources, as it is so big already and not to mention the rapid pace this industry is growing in. This means that it is important to consider how the data and the possible results can be used in the best way possible. This is not only in regards to the data itself, but also in how the user of the analytic system can use the data in the most efficient way.

As mentioned earlier, there is a multitude of data and data sources, everything from the video games themselves to the actual video game players. Therefore, it is also important to consider the possible outcomes of our research, and in addition to it we might figure out how and what we can use our findings for. There are several steps to take here, and a lot of different paths that can be followed. For instance, we want to figure out if video games are similar to each other, and possibly also how similar they are.

Chapter 3 Approach

This chapter will contain information about the approach to the problem statement, and what types of analysis techniques and methods that will be used moving forward. The first step in this process is to establish what the desired result are and what kind of accomplishments can be done later on. This means that it is not sufficient to determine whether or not games are similar or actually how similar they are, but in addition it could be beneficial to establish future potential uses of the results.

In order to determine if video games are similar or not, we must decide which kind of properties and patterns should we look for. In addition to this, we might want to figure out more about how similar they are, and potentially also what makes them related. In regards to the dataset that will be used moving forward, we know that it is not a complete dataset and that there are some missing data. It is unknown to us why the dataset has some gaps, and therefore we have to figure out what position we want to take in regards to it. The dataset consists of several different aspects in regards to the different video games that are represented. The most interesting part might be the timeseries of player count, with five minutes intervals for the top 1000 applications and one hour intervals for the next 1000 games [35].

3.1 Initial approach to our research

A sensible approach to begin with is to start with the player count for the applications that show one hour intervals. The reason for this is that the numbers are as mentioned in one hour intervals, which means that there are less data to work with. This means that the calculations and actual analysis of the data will be easier to conduct and simultaneously it will be less time-consuming. As a result of starting on this end, it is possible to start working on the data and do further analysis while the process is being started on the other part which contains more data as the intervals are shorter.

Initially, a theoretical foundation has to be prepared to be used further in our research. This theory will be based on our assumptions related to video games found in Steam and the specific patterns that these video games can provide. This theoretical statement has to be our foundation, and provide possibilities for further research. When a theoretical statement is in place, it will help us study our assumptions as something separate from the calculation itself.

3.2 Correlation coefficients and two-dimensional arrays

The general computational approach is to create a two-dimensional array in order to determine the amount of correlation between the video games, and in the event that there is a strong correlation, how big or small is it? There are several different ways and methods to measure correlation, amongst them are Pearson correlation, Kendall rank correlation, Spearman correlation and the Point-Biserial correlation. These methods have different approaches and it is beneficial to use more than one as they are different and therefore the results may vary [36].

Spearman's rank correlation coefficient or Spearman's Rho, is comparable to Pearsons correlation coefficient because it measures the strength of association between two variables. Spearman's Rho does not require continuous-level data, for the reason that it uses ranks instead of assumptions about the two variables, nor does it assume that the variables are distributed normally [37]. Because of this, we need to monitor the analysis and examine if something behaves the same way or not. In addition to this method we also need to use the correlation coefficient that will give us the numerical values between 1 and -1. These methods will both be used, and the reason for this is because they highlight different aspects. There are existing Python libraries for these methods among them are NumPy and SciPy. As mentioned we need to create a two-dimensional array with the different applications, and calculate all the correlations from it. Each application will cross itself one time in the array, and therefore the correlation at that point will always be 1, which is a perfect correlation.

After constructing the array, which is something that needs to be done multiple times, we will conduct further analysis on the array using clustering algorithms. Cluster analysis or clustering algorithms is the process of categorizing items based on their similarities. Clustering is something that is used in a variety of technological fields, for instance machine learning, bioinformatics, pattern recognition, among others [38].

3.3 Principal Component Analysis

We will begin with Principal Component Analysis (PCA) which “is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set” [39]. PCA cannot provide the accuracy that we need on its own, but when used in combination with a correlation coefficient it will provide more answers.

According to Begnum and Burgess (2004), when data mining is used to detect anomalies in computer systems, it is a comprehensive method where the outcome is uncertain. This relates to the fact that we do not know in advance if PCA will give us the desired result, but it is still a method where the goal is to select the axes that reflect the underlying symmetry features of an area in terms of the original parameterization [40]. Symmetry and similarities in different patterns is what we are looking for as the results of our initial analysis, in order to give us some information that we can use to support our hypotheses.

Clustering is a complex process, therefore there is several different algorithms for it. Aravind CR (2012) states that “Clusters found by one clustering algorithm will definitely be different from clusters found by a different algorithm” [38]. Therefore, it might be necessary to use other clustering algorithms alongside PCA. PCA is only a first step, which will give us a starting point and possibly also clarify how to move forward from there.

There is also several different factors that needs to be taken into consideration in this project. As mention, we will begin with one clustering method and move forward from

that. It might be possible to run PCA first and then conduct other types of analysis, but this is also dependent on several different things. It can be too prematurely to compare results from PCA with another clustering method, however, we do not have enough knowledge about this, which means that multiple algorithms has to be tested. In addition to this, we need to consider the fact that similarity can change over time, or that some video games are similar for a certain period of time. Therefore, it can be a good idea to monitor the clustering algorithms and the results they provide.

Another possible discovery might be that the video games that are the most similar also are under constant flux. At this point in time we do not have the correct knowledge about this, but we know that there is a possibility for that to be the outcome in some cases. For this reason, it is important to have more knowledge so that we are able to find the right path of analysis. It is beneficial to begin with one algorithm and use that on multiple correlation coefficients methods, like Spearman's Rho and Pearson R, rather than using multiple methods on one set of data.

3.4 Interpretations of results

After different examinations there are multiple ways to interpret the results, and possibly also different ways to use our results. The player count in our dataset is from December 14th 2017 to August 12th 2020. There is a reason to assume that minimum one of the video games in the dataset has been released during this time span. Because of that, it can be interesting to look at patterns in the newly released game, both on and surrounding the release date. In addition to the newly released game, we can investigate patterns within games that we assume is similar because of genre. When these potential patterns are compared, we might be able to detect whether or not some games have a decrease in player count during the time period in question.

One particular game that was released during the timespan of the dataset, is PlayerUnknown's Battlegrounds. It was first released as a beta version in March 2017, which gained a lot of good promotion for the game. Further, the video game was released in its full version on December 21st 2017, and we already know that this

game in particular is quite popular. This video game has been popular from its release date, and is still one of the most played games on Steam. Therefore, this might be a good example to investigate further, and look at if the increase in popularity within this game has affected any patterns elsewhere.

In addition to this it is potentially interesting to look at games that we know in general are very popular, as of May 14th 2013 the top five games on Steam are Counter-Strike: Global Offensive, Dota 2, Apex Legends, PlayerUnknown's Battlegrounds and Grand Theft Auto V [41]. These games are not all represented in our dataset, as New World were released in September 2021, but it presents an opportunity to access more data in addition to our dataset. The player counts here are not the results that we are looking for, but they can work as a pointer for us in our research.

In the case that there is a completely new game, is it possible to use the history from an older game within the same genre? This can also be helpful for new businesses that are in the process of developing a game. It is not necessary to invent the wheel by observing itself, when there is potential to gain more knowledge by studying games that appear to be similar. This potential approach can contribute in a way that new video games will have some kind of history, which can results in an opportunity to even out the playing field for new games, companies or services.

We do not know whether or not the data actually has the answer to our assumptions in regards to history of older games, but is still an important path that could be beneficial to explore. In order to find a possible solution for this particular question, we might have to collect more recent data in the background and compare it to what the results that we have gathered.

We will also have a selection of different games where each of them represent something in particular to us. Furthermore we will monitor these games more closely and see how where they turn out in our analysis and if we can use our findings for further research. These games could be of different genres in general, or we could look at first-person shooters or third-person shooters.

3.5 Alternative approaches

There are potential weaknesses in our approach, amongst them are the fact that it might be important to use multiple clustering algorithms. At this point we do not know whether or not it is a weakness or what PCA will do for us, so it is just a starting point. It is potentially also too premature to compare patterns with only one method, but this is something that we will have to research further after a while. It is beneficial to have one starting point in particular as this will help us to figure out what is potentially missing and how we can move forward from that.

Another potential issue is the fact that we do not know what happens if the similarity changes, if something is similar only for short periods or if something changes over a longer course of time. This is also something that will be more clear after working with our methods. In addition to this, we also know that our dataset has some gaps, and that is also a factor that is unknown to us. We do not have any information on the cause of the gaps, or potentially how these gaps can affect our research and results moving forward.

This research can be conducted through multiple different approaches, that in its own ways can be valuable in different ways. Our approach to similarity could be based on metadata about the different video games. This is something that we have available in the dataset, for instance genres, game developers, price model or publisher. For the purpose of research, this is definitely something that could be valuable in regards to similarity. Our aim is to attack this from a standpoint of having as little information as possible. The intriguing aspect of this is to see how far we could get with this as a starting point.

Additionally it could be interesting to collect our own data about video games. This would provide more control over the particular video games that are monitored, and it would be possible to select different games that we want to investigate further. This could be games that we know are popular amongst gamers, and possibly also new games that we know there is or have been a lot of anticipation and expectations surrounding the release of the game. This could possibly also open up an opportunity

to collect data from some other game distributors like Blizzard Entertainment or Electronic Arts. This way, we would have even more games to look for similarities within as these big video game distributors often have games that we know from experience are quite similar and can be seen as a competitor for some of the games found in Steam. This would require a more complex approach when it comes to data collection.

With an approach like this we would also be able to control the frequency of the data collection, for instance it could be interesting to see the player counts in one minute intervals as compared to five minutes. In this scenario it could also be possible to collect additional data, for instance about people promoting the video games, like a streamer or a company within the technological field. However, this approach comes with certain risks and would be very time-consuming in regards to how long it takes to collect the desired data, as well as knowing whether the data that is being collected is correct. This approach is something that could be used in future research, where it would be possible to investigate brand new games in the market and if they align with the patterns potentially uncovered in this project.

Another possibility would be to use something different from PCA to find key players. Begnum & Burgess (2007) have used the EventBank algorithm to identify key players in relation to anomaly detection. They state that “the basic assumption is that regular and periodic usage of a system will yield patterns of events that can be learned by datamining”. Generally, importance or interest ranking is based on statistical frequency analysis of classified symbolic events, whilst numerical measures have to be digitised and classified into elements for it to be possible to determine policies for further responses [42]. These kinds of methods are based on a history of events, and the ones that participated in them. The dataset could have been converted into events rather than numerical values, for instance top, increase, plateau and others. From there it could be possible to use EventRank or something similar to detect if video games have the same events simultaneously and use these result as a definition of similarity. The advantage of an approach like this, is that it would reduce the dataset considerably because the data would be simplified. In any case, our assessment at this point in time is that it would not be useful with a method that simplifies the data to begin the research.

There are several different ways to look at video games and patterns in them, which has different purposes. Floros and Siomos published an article in 2012, with the title "*Patterns of Choices on Video Game Genres and Internet Addiction*". In their paper they have attempted to identify choices made in regards to both online and offline games, to find meaningful predictors of Internet addiction [43]. This is a different approach from our research of similarity in video games, but it is an important approach to gain more knowledge about an important aspect of technology, internet and video games in general. In addition to this, Björk and Holopainen wrote a conference paper in 2003 about Game Design Patterns [44]. This is more about the video games in general, and the technical components of a video game.

Video games are a medium, comparable to tv shows, movies and books, which comes from different publishers, authors and genres. Therefore, just like the mentioned mediums, it is common that a person might prefer a certain genre, developer or author when it comes to the selection of what one might want to play next or what kind of new releases one individual is more excited about. In relation to this, there is a lot of adaptations between all this mediums. For instance, over the year a lot of books have been adapted into tv shows and movies, and lately the same has been done with video games as well. The most recent example of this is the video game *The Last of Us* which have been adapted by HBO into a tv show, and have received a lot of good appraisal and can be considered "the best video game adaptation of all time" [45]. This is something that is important for the video game industry as well, and can contribute to video gaming being more universally accepted.

A psychological aspect to video games, have been examined by Ryan et al. (2006) through the self-determination theory (SDT). SDT predicts that a player will enjoy a specific video game based on their psychologic need for autonomy, which is a sense of control, competence, which is a sense of performing well, and lastly relatedness, which is friends and relationships. Across all players, it was found that the subjective experience of the mentioned psychological needs during gaming made video games more motivating and appealing for a gamer. Furthermore, it shown that well-reviewed games, for instance The Legend of Zelda: Ocarina of Time, better satisfied needs

than what games that are considered fiascos did, for instance, a A Bug's Life. Different players found the same successful games to be satisfying in a different way for their SDT needs, and therefore, enjoyable in a different way. As this phenomenon suggests, preferences for an individual might moderate whether a particular video game satisfied or stifled SDT needs, and accordingly which gamers that will enjoy a specific video game [46].

A thought-provoking fact in regard to this is the potential variety of video games that we might find similar when performing our research. For instance, we might find some games that are similar because of the player count, and some of these video games that have high correlation and appear to be similar, maybe both critically well-reviewed video games as well as video games that are considered fiascos with bad reviews as mentioned above. This is not something that is directly related to similarity, but an interesting theory related to video games in general and might be an important subject to note for potential further research, and in the future it can possibly provide better insight when it comes to similarity and the factors that potentially can affect similarity itself.

Chapter 4 Result

4.1 Theoretical Model

This chapter will present our theoretical basis surrounding the degree of similarity in video games from Steam, we will first present a list of assumptions and then move forward to some project hypotheses for further investigation.

4.1.1 Initial assumptions

A1: *All games **g** have a periodic pattern **p** which holds for a period **t***

This states that all games found in the dataset have a pattern, and that this particular pattern holds for a period of time. If ***t*** is high, it tells us that there is a more predictable pattern over a longer time. On the other hand, if ***t*** is low it tells us that the pattern is more dynamic and can change. The actual length of ***t*** is unknown to us, but it is still important as patterns means that there is a variety in the amount of players. This is something that will likely have a waveform, day after day. It is important to notice that these patterns will not be coherent, as there is a difference both when it comes to daytime and evening, and weekdays and weekends.

On a general basis, we know that a lot of people are busy during the daytime, this could be with work, at school or studying. This means that a video game is more likely to have more active players during the evening, at least from Monday through Friday. At the same time, we know that most people that have a daytime job or go to school probably have more free time on the weekends and therefore there might be more players from Friday evening throughout Sunday.

As well as having more time off from work and other obligations during the weekends, there is also the possibility to be more social and for some that means playing video games. For instance, some people arrange a Local area network (LAN) party during weekends, both in a smaller scale at home with friends and there are

larger events with more people and happenings. This can also be seen in relation to holidays, where students especially might have more time. During Easter in Norway, there is a LAN party hosted called The Gathering, which had approximately 5000 participants in 2018 [47].

A video game, for example *Counter Strike: Global Offensive* would start every day at 12 am with an amount of active players, and reach its highest number of active players during the course of twenty-four hours sometime during the night. According to SteamDB [39] on May 14th 2023, Counter-Strike: Global Offensive's twenty-four hour peak is 1 759 286 players, and it's all-time peak is approximately 1.8 million players.

> **A2:** *Every pattern **p** has a similarity **s** with every other pattern which can be established numerically*

This statement is a method for us to establish similarity, and the potential similarity comes in a numerical value. Our interest is in the degree of similarity, not how similar they are or what makes them similar. For instance, if we look at three different games that has a different pattern, what is the distance between these?

The important aspect here is not the pattern of one selected game that is interesting for the fact that one pattern in itself will not explain anything in particular. The desired approach is to look at the similarity between two patterns. It is not meaningful to us whether the pattern of one game has an abnormal increase or decrease, as long as the two video games in question have the same increase or decrease. For instance, we could look further into the patterns of Tekken 7 and Mortal Kombat 11, which are both found in Steam. These two are both fighting games, and can both be played as single player or multiplayer. Therefore, it is possible that these video games have a similar pattern.

A3: *For patterns **p** and **p'**, the similarity function **s**(p,p') would return a numerical value describing their similarity*

If we look at three arbitrary games, we can examine the distance between these. Are two of this more alike each other, are all of them similar or are they different from each other? If we learn that two of them are similar, it might be possible to apply a transitive relation.

A relation is transitive if [48]:

$$(a, b) \in R \ \& \ (b, c) \in R, then \ (a, c) \in R$$

This means that if a is similar to b, and b is similar to c, then a and c are also similar. For instance, if we look at the games Guild Wars, The Elder Scrolls and Lord of the Rings Online, we know that these three video games are massively multiplayer online role-playing games (MMORPG) and therefore the relation between them might be transitive.

We cannot be sure if a relation is transitive or not, because there are multiple factors involved. This is something that needs to be investigated further in order to determine transitivity. For instance, you could have three different video games, game ***a***, game ***b*** and game ***c***. Game ***a*** can be similar to ***b***, for example with having the same genre like real-time strategy (RTS) or first-person shooter (FPS), whilst game ***b*** and game ***c*** might have the same theme, like Vikings or military. Based on this hypothetical example, it can appear that there is a transitive relation in this situation but when we investigate further we learn that we cannot be sure of the relation actually is transitive.

A4: *Based on this similarity **s**, groups of games can be established*

This is a statement that we assume is theoretically possible. If our initial use of PCA fails, that does not mean that we have to reject this clustering algorithm at once.

At the same time, if we are able to identify clusters of video games, that does not imply that the results are verified or correct.

In any case, PCA will provide clusters after its applied, but we need to consider whether or not these groups are meaningful. We do not have an approach to verify the clusters that are provided, and there is no clear-cut solution this. With a general knowledge of video games and particularly video games that are well-known, we have our own intuition in regards to the aspects we already know. Although, this is not a robust solution to the problem.

There is a need for a more critical review, combined with other assumptions about the different video games. We can assume that video games in the same genre will be in the same cluster, as it is implied that the same genre equals similarity. For instance, we have PlayerUnknown's Battlegrounds, Hunt Showdown and Day Z these are all battle royale games found in Steam. The term battle royal has its origin from a Japanese cult film with the same name as the genre, and the characters in the film has to fight until their death. A battle royale video game is a multiplayer game with around 100 participants, where the goal is to be the last man standing [49]. Based on what we know about these games it may be correct to assume that they are in fact, similar. This is only an assumption, and there needs to be further research in order to gain more knowledge. If the outcome does not match our assumptions or what our intuition tells us, it does not necessarily mean that the method is wrong, it will just be a surprising result.

It is not necessary for a group of video games to have make sense to us, that means that it does not have to have any common denominators like whether or not it is a first-person shooter or a platform game. On the other hand, it would still be interesting if there is a meaning or a pattern when it comes to the group itself. When the analysis is conducted, we can, with the help of correlation look further into how similar the most similar video games are.

If we carry out multiple rounds with a varying period $\boldsymbol{t}$, meaning different time periods in the dataset. For the reason that we have data from more than 2.5. years, so we can look into both different days, different times of day, weekends or holidays. This

will help us to illustrate how robust the similarities are, because we will have more data and therefore more knowledge.

It could also possible to gather more information about different events like LAN party's and other gaming related happenings, and explore what happens to the patterns at this particular time. Other gaming related happenings could a stream, which means to broadcast the screen of a video game, in most cases with a video of the player, to a live audience [50]. There can be communities or well-known streamers that has a stream for a special occasion, and might even do some games with their viewers. It would probably not be possible to see any changes in patterns if a streamer with low ratings and an average number of followers. However, there are different streamers that have multiple millions of followers, for instance Ninja which as of May 14th 2013 with more than 18 million followers [51]. This is a number that definitely can have impact on player count.

4.1.2 Project hypotheses for further investigation

Furthermore, we will formulate some hypotheses. These are assumptions, that can be proven false after they are tested.

H1: *The similarity **s** of two patterns change with different values of **t***

This hypothesis tells us that if we have two patterns with similarity, and the period of time changes, the patterns themselves and the similarity will change as well. Even though our initial research might show similarity in two patterns, there can be different aspects to them. Our two similar patterns can never be exactly the same, as that would mean that we are looking at the same game. Two patterns might be similar for a period of time, and different for another. Which means that it is correct to assume that the similarity will change if the time period changes.

As mentioned, it is possible to look at different time periods during a particular day, this is something that we from our intuition and previous assumptions think that will change from day to night. It is also possible to look at longer periods of time, for

example the first six months of a year and the last six months. There are many combinations here that can be used to learn more about the patterns.

We also need to consider whether or not it is fair to compare different periods of time, as there is potential to retrieve more knowledge about games so that our analysis will be more robust. This is not something that has to be taken into consideration at this point, but may be a potential development in the future.

H2: *Groups of similar games change with different values of* ***t***

When PCA is conducted, we know that it will give us groups of games, but we do not know anything about the size of groups or what kind of video games they will contain. If we look at different periods of time, it is possible that the group itself can change. This is because games can be similar for one period, but different in another period. The potential results here may vary, and it depends on what the time period actually is. It could be hours, days, weeks or months that are being compared.

We will not know what the actual reasons for the changes in similarity, but this could also be in relation to release dates of new games or new downloadable content (DLC) for a particular game. We know from experience that some games have been popular and continue to be so for years and years, and on the other hand there could be new games that are popular as they are released. This popularity can be consistent for weeks or months, but if new games does not hold the same standard as the classic games like Counter-Strike: Global Offensive or PlayerUnknown's Battlegrounds, many people will go back to games like these.

H3: *A high similarity over a long* ***t*** *does not guarantee a high similarity for a smaller* ***t'*** *within the original* ***t***

If we look at a group of games or two different patterns, for instance through the course of 3 months, or more. If our analysis then tell us that these two patterns are similar through this period of time, it is important to look further into this. The reason for this is that even though something is similar for one period, it does not

mean that it will be in smaller periods within the 3 months. Similarity can change in several different ways.

By looking further in to the timeline that we have chosen, we can investigate different blocks of time in the timeline. Two video games can be similar for three months, but the similarity here can change throughout the timeline. If we then look closer at weeks or days within the 3 month timespan, the video games might not be as similar all the time. It is possible that there are periods within the 3 months that are similar as well, but similarity can change and we are interested in knowing if similarity is something that changes throughout the course of a particular timeline.

These hypotheses will be tested further in our research, in order to gain more knowledge about the similarity. This means that we will determine whether or hypotheses are true or false. In regards to the assumptions that we have made, this is not something that we will test in practice, but we will still consider whether or not these assumptions have been useful to us.

4.2 Results

The first script in the research was developed in order to calculate correlation between two files in the dataset. In order to do some testing in regards to this we chose twelve different video games, to contribute to getting a better overview of how the results looked. The games were chosen based on games that were assumed to be similar, with some games that were assumed to be different from each other.

The twelve games that were chosen are the following:

- Total War: Warhammer II
- Cities: Skylines
- Sid Meier's Civilization IV
- Dead Island: Riptide
- Dying Light
- Payday 2
- Tom Clancy's Rainbow Six Siege
- Counter-Strike: Global Offensive
- Counter-Strike
- Football Manager 2017
- NBA 2k18
- Final Fantasy XIV

Total War: Warhammer II (TWW2) is a turn-based/real-time strategy (RTS) game, published by SEGA. The game is set in the world of Warhammer Fantasy world created by Games Workshop. The game is a sequel to Total War: Warhammer, and was followed by Total War: Warhammer III which was released in 2022 and therefore, is not a part of the dataset used for conducting this research [52].

Cities: Skylines is a game revolving around city building, and it is a single-player game that is open-ended. Players can undertake urban planning by controlling zoning, road placement, public services and public transportation of an area. There are various elements of a city that is managed by the player, including budget,

health, traffic, employment and pollution levels [53].

Sid Meier's Civilization IV is the fourth instalment in the Civilization series. This game is a turn-based strategy game, where the main objective is to construct a civilization from limited resources [54].

Dead Island: Riptide is zombie themed video game and an action role-playing game (RPG), and it is the sequel to the Dead Island game which was released in 2011. The game is played from a first person point of view, but it is not a first person shooter because of the focus on melee combat. Melee combat is hand-to-hand combat in battles at a close range [55].

Dying Light is an open world first-person survival horror video game, published by Warner Bros Interactive Entertainment. The game is a successor of the Dead Island series. Similar to Dead Island, Dying Light is also played in a first person point of view [56].

Payday 2 is a co-operative first-person shooter action game developed by Overkill Software. Up to four players can cooperate in a heist in this game, and the game features twelve heists, where some of them take place over multiple dates and locations. These heists include robberies of banks and shops, as well as production and distribution of narcotics [57].

Tom Clancy's Rainbow Six Siege (TCRS) is a first person shooter, and an online tactical video game. The game revolves around environmental destruction and cooperation between players. The player can both play as attackers or defenders in this game [58].

Counter-Strike: Global Offensive (CS:GO) is a team-based first person shooter, developed by Valve Corporation, which is also the developer of Steam. Players compete in multiplayer matches using different weapons and tactics [59].

Counter-Strike is just like CS:GO, a team-based first person shooter. The game was originally released as a mod for Half-Life [60]. A mod is video game modification,

which is an adaption of a video game that is often based on the engine behind the game, but elements in the game is changed [61].

Football Manager 2017 is a part of the Football Manager franchise, and is a football management simulation video game. The game involves taking charge of a professional association football team, playing as the manager. This role involves signing players to contracts, manage finances and talk to the players [62].

NBA 2K18 is a part of the NBA 2K franchise, and is a basketball simulation game. This game simulates the National Basketball Association (NBA). There are different modes of this game, where you can play as a team manager, or as a player where you play through the career of their own player [63].

Final Fantasy XIV is a massively multiplayer online role-playing game (MMORPG). The game features a persistent world where players can interact with each other and the environment. When starting this game players can customize their characters, and choose a game server for the character to exist on [64].

These games were chosen based on prior knowledge of video games, and this step was conducted both to ensure that the script was working and to examine whether or not the initial assumptions were somewhat correct. The twelve games were chosen for several reasons, in regards to the genre of the video game and how the video game is played.

The first group in the selection is *Total War: Warhammer II, Cities: Skylines* and *Sid Meier's Civilization IV* because these games revolve around world building and is based on strategy, both turn-based strategy and real-time strategy.

The second group in the selection is *Dead Island: Riptide* and *Dying Light*, where the common denominator is that they are zombie games based on survival.

The third group consists of *Payday 2* and *Tom Clancy's Rainbow Six Siege*, because these are based on cooperation and strategy.

The fourth group is *Counter-Strike* and *Counter-Strike: Global Offensive*, where *Counter-Strike* is the first video game released in the Counter-Strike series, and *Counter-Strike: Global Offensive* is the last. These were selected because of the aspect of doing research on a predecessor and successor.

The fifth and final group is *Football Manager 2017*, *NBA 2k18* and *Final Fantasy XIV*. These games were selected as alternative games, because their genre differs from the other video games in the selection. Football Manager 2017 and NBA 2k18 are assumed to be similar as they both are sports games, but are in this group because they appear quite different from all the former mentioned video games in the selection.

In general, all of these games can be compared to each other, both when it comes to assumed similarity, and assumed dissimilarity. The reason for this is that even though the majority of the twelve selected games have the same genre, their subgenre often differs and have a broader spectre of categories.

	Genre	Subgenre	Perspective	Multiplayer	Online	Free-to-play
Total War: Warhammer II	Action	RTS	Top-down	Yes	Yes	No
Cities: Skylines	Simulation	Strategy Sandbox	Top-down	No	No	No
Sid Meier's Civilization IV	Strategy	TBS	Top-down	Yes	Yes	No
Dead Island: Riptide	Action	Survival horror	First person	Yes	Yes	No
Dying Light	Action	Survival horror	First person	Yes	Yes	No
Payday 2	Action	RPG	First person	Yes	Yes	No
Tom Clancy's Rainbow Six Siege	Action	TFS	First person	Yes	Yes	No
Counter-Strike: Global Offensive	Action	TFS	First person	Yes	Yes	Yes
Counter-Strike	Action	TFS	First person	Yes	Yes	No
Football Manager 2017	Simulation	Sports	Simulation	No	Yes	No
NBA 2K18	Simulation	Sports	Top-down	Yes	Yes	No
Final Fantasy XIV	MMO	RPG	Third person	Yes	Yes	No

Table 4.1: The 12 selected video games with some additional information in the form of values related to each video game

Table 4.1 presents the 12 selected video games for the narrow approach, which will be explained later on, and a selection of values related to these games, in order to categorize them and some of their features. This also provides some general information about the games, and factors that apply to why they were selected. Several of them have the same genre, but subgenres differ for the majority of them, and the selection were made both because of assumptions of similarity with some features that are assumed to be quite different from the others.

All these games have been compared and calculated the correlation coefficient for, which is presented in the Table 4.3. Testing was also conducted by calculating correlation on the games against itself as well, to make sure that we got the expected result which is a perfect correlation and resulting in the number 1. Results from a

selection of these games will be presented more in depth later on, with the data split into intervals.

4.2.1 How is similarity measured?

Into our approach of research when it comes to discovering whether or not there is any similarities in these video games, there is several different approaches that could have been chosen for conducting this research. There has been a big question of how to approach this, because the dataset contains a huge amount of data. The approaches that can be made is, for instance, in a broad spectre where we go into it with no prejudiced thoughts or opinions, which leads to no bias when it comes to the results. This also means that we will produce more data, which gives us more tools to work with when it comes to forming a conclusion or opinion. Although, as there is an extensive amount of data, this is an approach that could possibly affect the quality of life within this research.

Another approach is to go into the research more in depth, with certain assumptions based on our own intuition and opinions. This is possible because we already have a knowledge of video games, and therefore we will be able to make assumptions when it comes to video games being similar or not. This requires some testing to see if our initial thoughts are correct or not, and can be done by choosing three video games that are represented in the dataset. With these three games, the point is to choose two games that we presume will show themselves to be similar, and a third game that is different from the other two in order to determine if this is correct or not. If the results of this is something other than what we expect, it does not mean that the approach or results itself are wrong, but it helps provide better insight when it comes to the question of similarity.

These two methods are important in their own way, as they answer different questions and provide further knowledge in two different ways. It may seem like the best approach is to choose the first approach mentioned, that revolves around a broad spectre, as this is the approach that would provide the most results to work with. Although, as mentioned the dataset is large, which leads to a lot of results that will not really tell us anything and the difficulty lies in what to actually do with these

results. Therefore, the method of approaching the dataset is both the broad and the narrow version. By conducting analysis and calculations on a small set of data to begin with, which is handpicked, we will learn more and create a better foundation moving forward.

4.2.2 Explanation of the steps when it comes to conducting our research

The first step of the process has been to calculate correlation between two different video games in the dataset. We also conducted a correlation analysis on the same file to make sure we ended up with a perfect correlation, where the expected result is the number 1, to make sure it is accurate. The reason behind this approach is to create smaller building blocks which can be used along with the others that are created, to give us more tools and a stronger foundation to corroborate the problem statement in this thesis.

The process was started with selecting three games from the dataset, two of which we assumed would have a positive correlation in order to corroborate the theories we have when it comes to which games are similar or not. In addition, a third game that we assume is different from the former two, and therefore, might result in a negative or low correlation. The three games selected for this initial part, was Total War: Warhammer II, Cities: Skylines and Dying Light. The correlation for the whole profile of these games are presented below.

	TWW2	**Cities: Skylines**	**Dying Light**
TWW2	1	0,611317	0,306886
Cities: Skylines	0,611317	1	0,477809
Dying Light	0,306886	0,477809	1

Table 4.2: Correlation coefficients from three games used in the initial testing when developing the scripts

As we can see in Table 4.2, Total War: Warhammer II and Cities: Skylines have a slightly higher correlation than Total War: Warhammer II and Dying Light, while the correlation between Cities: Skylines and Dying Light is at a point between these two other values. As mentioned, this is the correlation values from the whole profile itself, and therefore it might not tell us as much because it is such a large area of data.

The next step is to approach the assumptions that the similarity will change as a result of time. Considering that each file in this section of the dataset is presented by player count for every fifth minute of a day during a 973 day period. This means that there is 288 points during the time span of 24 hours. Therefore, the next calculation pertaining to similarity revolved around splitting the data into days. Moving forward from there, there have been made calculations for week by week, which is based on 2016 points during a week. Furthermore, calculations have also been made for approximately a month, where 8064 points was used, to represent four weeks as the amount of days during a month differs. Nevertheless, four weeks is a representative choice to provide results for a longer period of time.

Calculations for half a year, and a year have also been conducted. This is based on the introductory aspect of using 288 points for one day, where one year as we know it has 365 days, and therefore half a year is 182,5 days, which has been round down to 182. It is also important to note that these calculations will span from the beginning of the dataset, which is December 14th 2017, and a year moving forward from that. The dataset itself is as mentioned 973 days, which is slightly more than two and a half years in total. A different approach could have been a calendar year in itself, where we have data from all of 2018 and 2019, but this approach for results within a time period of a year is coherent with the previous elucidated starting point.

Additionally, all files have been calculated against each other, in a script created for total correlation. As there is 1000 video games in the first part of the dataset, and by calculating all of these against all the others, the outcome of this is extensive, and amounts to approximately half a million correlation coefficients. These results are overwhelming as there is so much data to process, and there lies a challenge in both how to handle this data and how to move forward with it. Nevertheless, these results can give us some additional knowledge about these video games, for instance which games prove to have the highest and lowest correlation result. As well as how many video games that prove to have a high correlation and how many that prove to have low correlation. We already know that a correlation coefficient ranges from the result of 1 to -1, but it is still difficult to determine which numbers that can be classified as high and which can be classified as low. These will still be interesting results that

provide further information which can be looked at in comparison with the initial assumptions made within this project.

The different experiments in this project are:

- Complete correlation
- Daily interval based correlation
- Weekly interval based correlation
- Monthly interval based correlation

Complete correlation is a calculation of the whole profile of two different video games, which gives us the correlation coefficient for all the data that one profile contains. As mentioned, these profiles contain data from every fifth minute over 973 days, which does not necessarily give us that much information. For instance, if there supposedly is a high correlation within parts of the dataset, it might not be visible when taking the whole profile into account. The reason for this is that there are a lot of information and some values can be erased over time. Some initial tests were conducted to make sure that this script worked and provided a result between 1 and -1. The first video games that were tested was Counter Strike and Killing Floor, where the correlation coefficient is 0,723556. Following these games, Supreme Commander: Forged Alliance and Oddworld: Abe's Oddysee were tested, where the correlation coefficient is -0,011967. Lastly, Stardew Valley and Minecraft were tested, which has a correlation coefficient of -0,123014.

Daily interval based correlation is a calculation based on each day in the dataset, and the 288 points that represent one day in the file for each video game. The player count is represented by numbers for every 5 minutes during the day, starting at 00:00, ending at 23:55. This gives us more information about the video game, and offers the opportunity to look at specific dates, and can potentially answer the question of how often we experience a similar day.

Weekly interval based correlation is a calculation based on the previously mentioned daily calculation, as there is 288 points during a day, that times 7 is 2016 which gives us representation of a week. As with the daily correlation this also provides

information about a smaller period of time, and can give us insight when it comes to specific weeks.

Monthly interval based correlation is also based on the two previously mentioned, with the basis that there are 288 points during one day. As the different months during a year consists of a different amount of days, this is set to 4 weeks as that is considered a well enough representation of a month. By doing this calculation we can look at smaller parts of the dataset, which is of course, longer than the previously mentioned intervals. Simultaneously, it gives us more insight into the video games and how the numbers change as the period of time changes.

As stated earlier 12 video games were chosen and categorized into different groups based on assumptions regarding these video games, the reasoning behind this is in regards to genre, subgenre, how the video game is played, the intensity the video game and the theme in general. After this 12 games were selected, they have all been compared to one another and all the correlation coefficients is listed in Table 4.3. The groups within this game does not mean that those groups will be presented more in depth, but the groups are provided as an explanation of the differences between them and therefore the in depth explanations will involve two video games from one group with the addition of one video game from another group.

	TWW2	Cities: Skylines	Civilization	Dying Light	Dead Island	Payday 2	TCRS	CS:GO	Counter Strike	Football Manager	NBA	Final Fantasy
TWW2	1	0,611	0,647	0,307	0,412	0,270	0,407	0,716	0,471	-0,189	-0,355	0,447
Cities: Skylines	0,611	1	0,732	0,478	0,503	0,190	0,678	0,759	0,595	-0,047	-0,265	0,402
Civilization	0,647	0,732	1	0,270	0,470	0,265	0,500	0,659	0,626	0,064	-0,299	0,556
Dying Light	0,307	0,478	0,280	1	0,315	0,221	0,555	0,498	0,341	0,004	0,034	0,168
Dead Island	0,412	0,503	0,470	0,315	1	0,269	0,433	0,467	0,504	0,173	-0,047	0,256
Payday 2	0,270	0,190	0,265	0,221	0,269	1	0,177	0,314	0,391	0,381	0,295	0,071
TCRS	0,407	0,678	0,500	0,555	0,433	0,177	1	0,705	0,552	0,063	-0,008	0,229
CS:GO	0,717	0,759	0,659	0,498	0,467	0,314958	0,705	1	0,732	-0,069	-0,328	0,412
Counter Strike	0,471	0,595	0,626	0,341	0,504	0,391656	0,552	0,732	1	0,524	0,164	0,129
Football Manger	-0,189	-0,047	0,064	0,004	0,173	0,381036	0,063	-0,069	0,524	1	0,748	-0,398
NBA	-0,355	-0,265	-0,299	0,034	-0,047	0,295579	-0,008	-0,328	0,164	0,748	1	-0.514
Final Fantasy	0,447	0,402	0,556	0,168	0,256	0,071912	0,229	0,412	0,129	-0,398	-0.514	1

Table 4.3: The 12 selected video games, with correlation values after being calculated against each other

Table 4.3 presents the correlation coefficient between the 12 selected games for the narrow approach in this project, and they have all been compared against one another, by using the former mentioned experiment the revolves around complete correlation. Further, we will look into some of these video games more in depth, by looking at daily, weekly and monthly correlation between these games, and additional numbers surrounding this.

Moving forward, there will be different distribution plots presented for the video games that will be presented more in dept. The numbers in these distribution plots are presented in hundreds, as opposed to decimals, this was a necessary change to these particular plots in order to present the numbers graphically. These presentations will be of 6 different video games, in two groups of three. The first group is NBA 2k18, Football Manager 2017 and Final Fantasy XIV, and the second group is Total War: Warhammer II, Sid Meier's Civilization IV and Payday 2.

4.2.3 Selected games: NBA 2K18, Football Manager 2017 and Final Fantasy XIV

In this selected group, we assume that NBA 2K18 and Football Manager 2017 are similar games, as the theme of these games are sports and revolves around the simulation of the practice of sports. Final Fantasy XIV on the other hand, is a MMORPG and a more intense game based on fantasy and adventure. A common denominator, is that all of the above are games that are a part of a franchise that has been active and releasing games over a long period of time. The first Final Fantasy video game was released in 1987, the first NBA 2K video game was released in 2002, and lastly, the first Football Manager game was released in 2004. Both NBA 2K and Football Manager games in the franchise are released annually. They all have both predecessors and successors, and possibly a loyal fanbase which continuously play these video games, both previously released versions as well as the new releases in the series.

The correlation on calculated on the whole profile of these three games are as follows:

- The correlation coefficient between NBA 2k18 and Football Manager 2017 is 0,74859
- The correlation coefficient between NBA 2k18 and Final Fantasy XIV is -0.514751
- The correlation coefficient between Football Manager 2017 and Final Fantasy XIV is -0,398289

As stated in the previous paragraph, the assumption was that NBA 2k18 and Football Manager 2017 are similar games, while Final Fantasy XIV are different from the two others. As we can see from the bullet list, the correlation between Football Manager 2017 and NBA 2k18 is positive and high, while Final Fantasy XIV has a negative correlation with both the others.

NBA 2K18 versus Final Fantasy XIV

In addition to calculating correlation on the whole profile, we have looked at different periods of time in order to get more in depth knowledge, and to see if something changes over time or not. The following diagram shows the distribution of the

correlation values for NBA 2K18 and Final Fantasy XIV on a daily basis, which are to video games that we assume are different to each other and therefore might have a negative correlation.

NBA 2K18 versus Final Fantasy XIV - daily interval based correlation

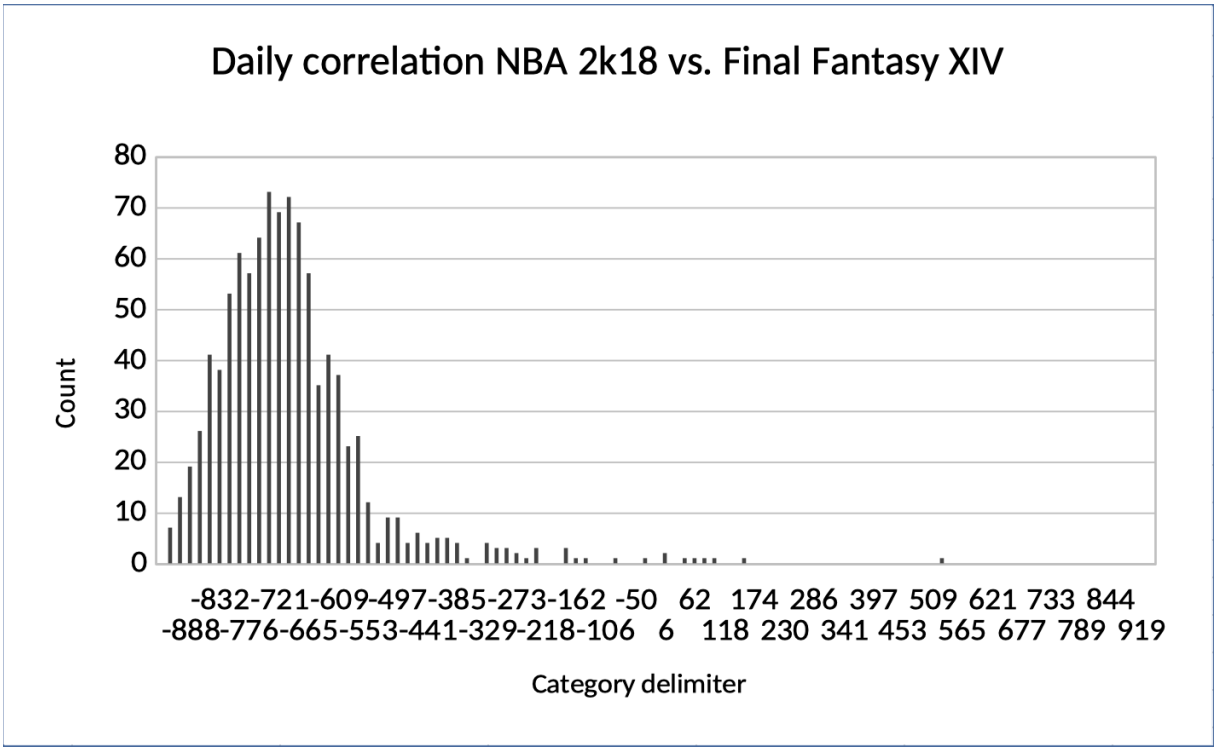

Figure 4.1:Daily correlation for NBA 2K18 versus Final Fantasy XIV, presented in a 1% frequency distribution plot

In Figure 4.1, the daily correlation values between NBA 2k18 and Final Fantasy XIV are presented. These numbers are normalized to hundreds, which is why they do not appear as decimals. This diagram shows us that approximately 99% of the correlation values for the 973 total days included in the files for the mentioned two video games, has a negative correlation value. The average correlation value between these two games are -0,6880015 and the median is -0,728982. As shown in this figure, the majority of the correlation values are high and negative, and a small part of them are closer to the correlation value of 0 as seen in the tail towards the right side of the figure.

NBA 2K18 versus Final Fantasy XIV - weekly interval based correlation

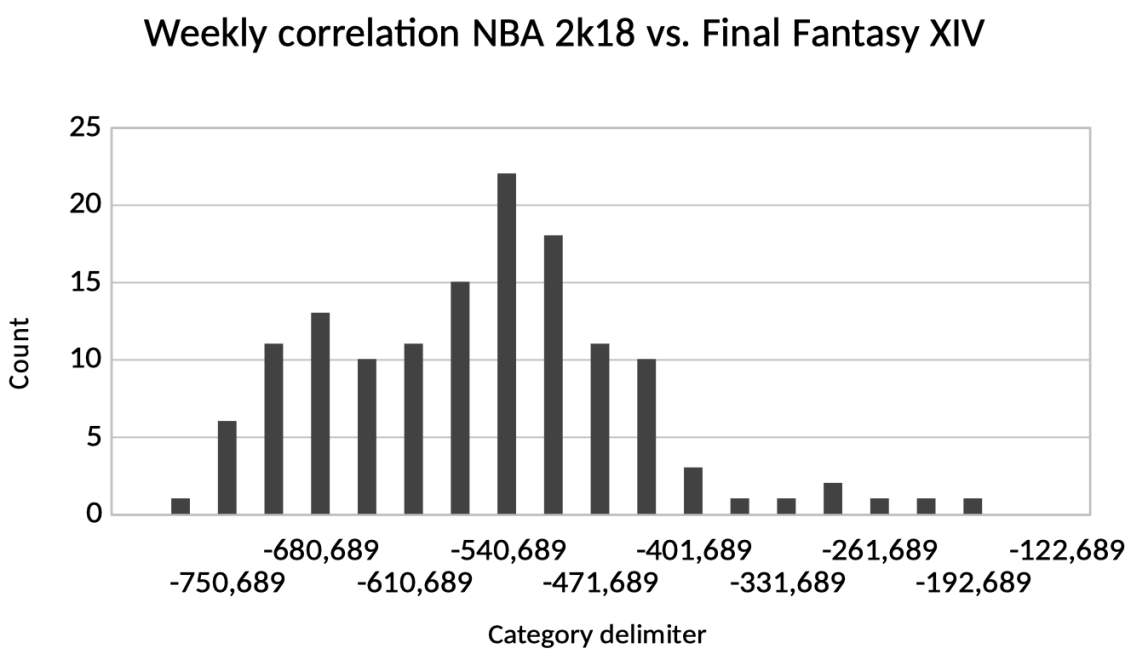

Figure 4.2: Weekly correlation for NBA 2K18 versus Final Fantasy XIV, presented in a 5% frequency distribution plot

The values in Figure 4.2 represent the same data as shown in Figure 4.1, but the correlation values have been measured over weeks instead of days. There are a total of 139 weeks within the dataset. As seen this figure follows the same pattern and curve as Figure 4.1, but the field moves a bit more towards right. This means that the values change in a small degree when the time interval is changed.

NBA 2K18 versus Final Fantasy XIV - monthly interval based correlation

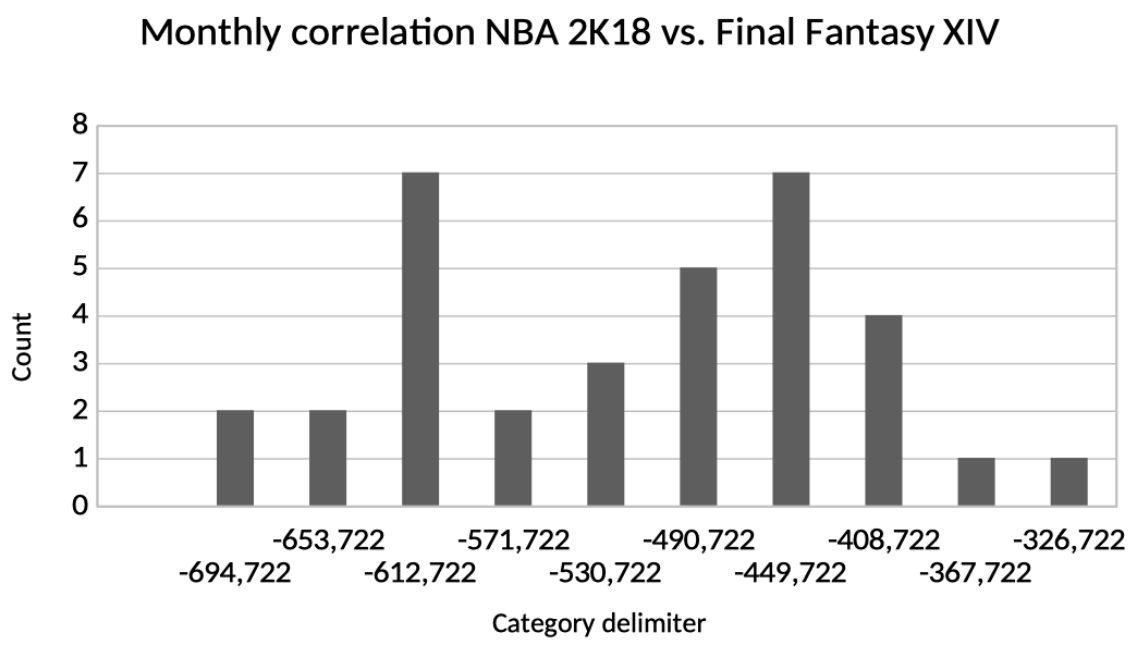

Figure 4.3: Monthly correlation for NBA 2K18 versus Final Fantasy XIV, presented in a 10% frequency distribution plot

Figure 4.3 presents the monthly intervals for NBA 2k18 versus Final Fantasy XIV, and compared to daily and weekly intervals, as presented in Figure 4.1 and 4.2 respectively. The difference between the results in this figure compared to the former two, is that the correlation values are a lot more spread. They are a bit lower as well, even though the correlations in this case are negative. The lowest correlation value in this specific interval is -0,326621, the highest is -0,734722, and the average is -0,53549.

NBA 2K18 versus Football Manager 2017

This section will deal with the correlations between NBA 2k18 and Football Manager 2017, which are two video games where the assumption is that they are similar based on the fact that they are both sports video games.

NBA 2K18 versus Football Manager 2017 - daily interval based correlation

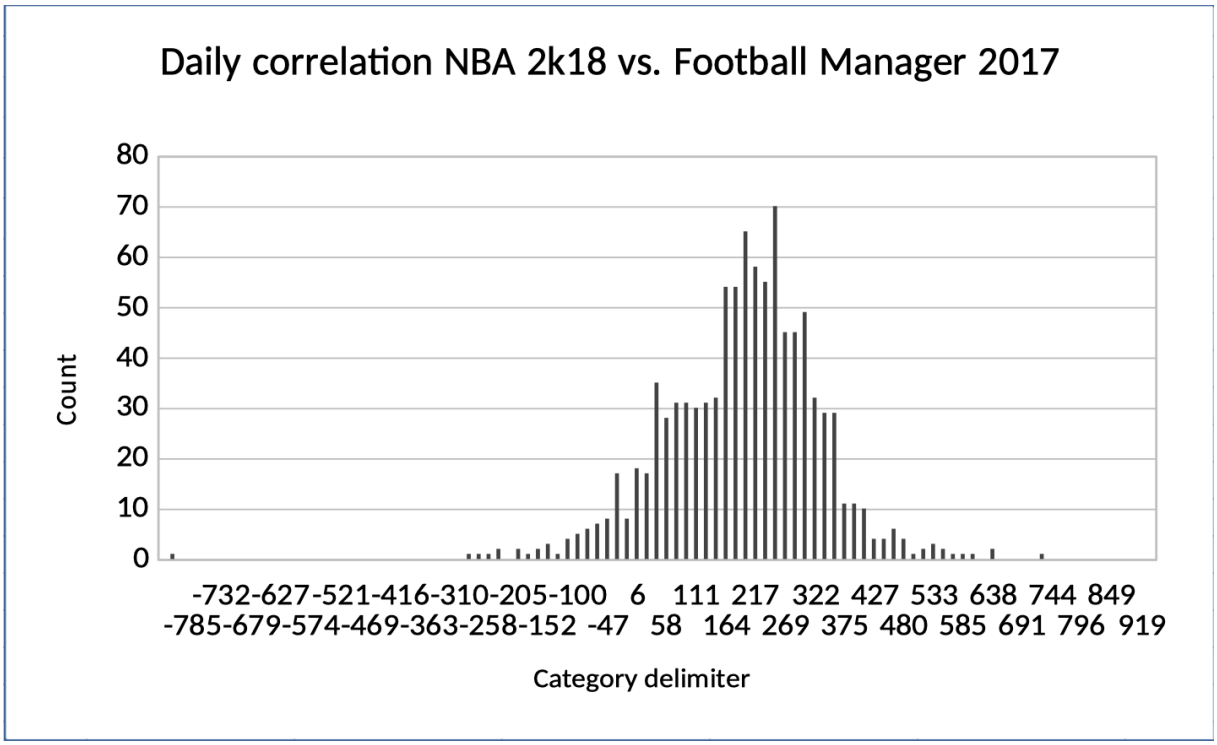

Figure 4.4: Daily correlation for NBA 2K18 versus Football Manager 2017, presented in a 1% frequency distribution plot

Figure 4.4 shows the correlation values between NBA 2k18 and Football Manager 2017, with the daily interval that consists of 973 days. These are two games that are assumed to be similar because they both revolve around sports simulation, as we can see from this figure, the values are mostly of positive correlation, but it is not very high and the average for the daily intervals are 0,185445.

NBA 2K18 versus Football Manager 2017 - weekly interval based correlation

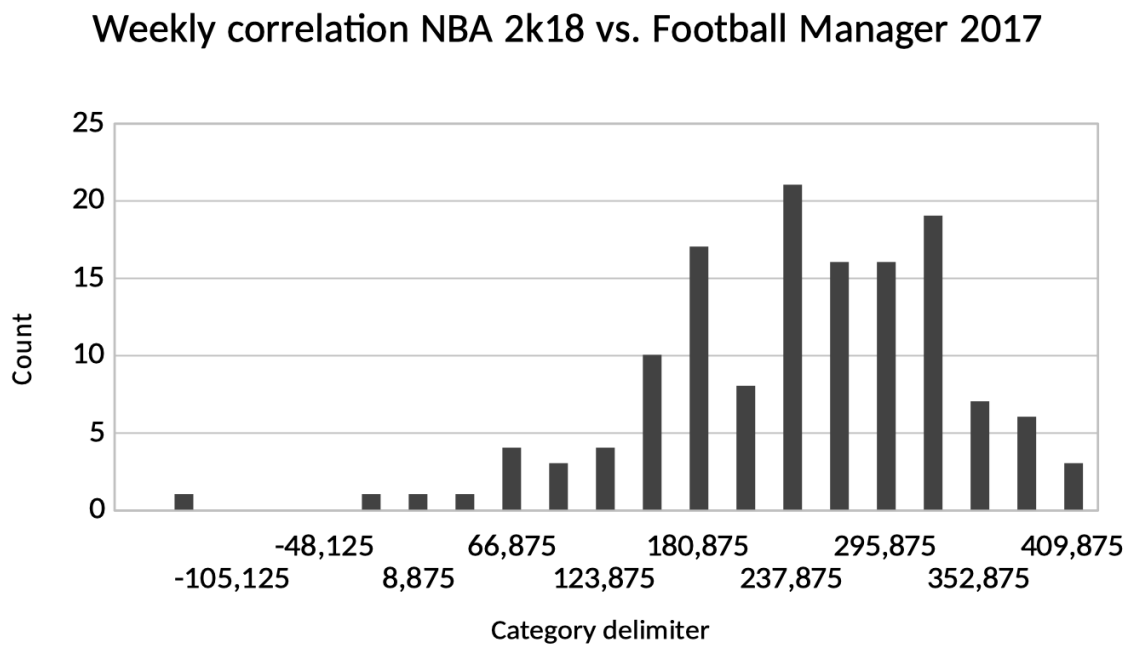

Figure 4.5: Weekly correlation for NBA 2K18 versus Football Manager 2017, presented in a 5% frequency distribution plot

The weekly interval between NBA 2k18 and Football Manager 2017 as shown in Figure 4.5, changes slightly from Figure 4.4, the numbers are not that different from the former figure, and the average in this case is 0,229398. The correlation coefficients have moved more towards the right, but the pattern of the figure is very similar to the daily correlation in Figure 4.4.

NBA 2K18 versus Football Manager 2017 - monthly interval based correlation

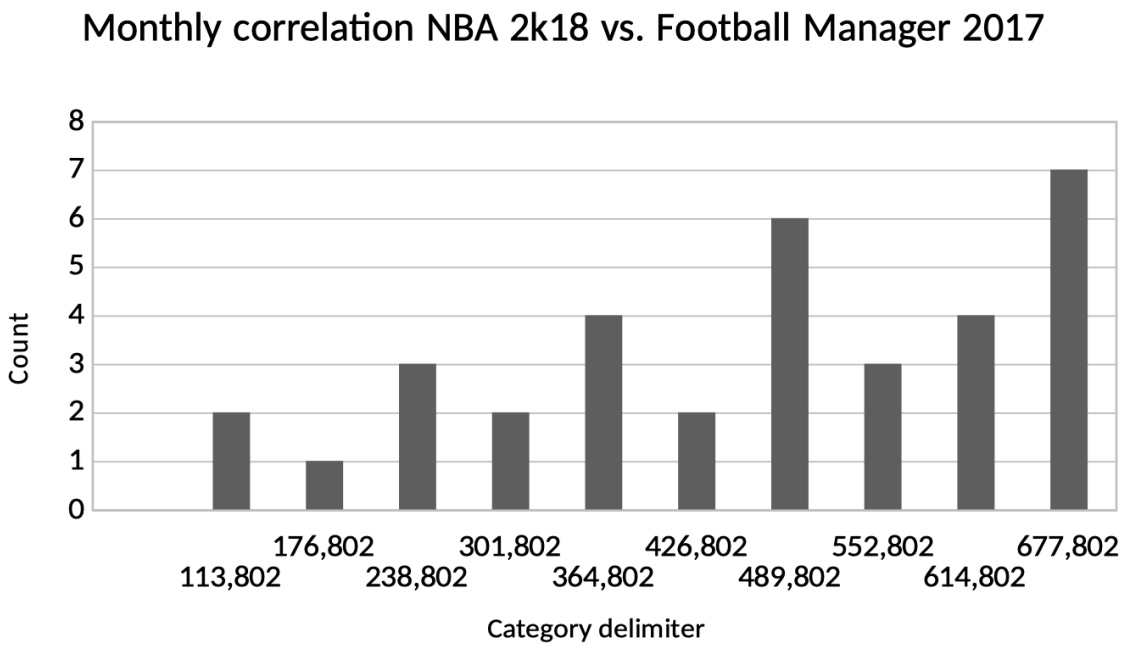

Figure 4.6: Monthly correlation for NBA 2K18 versus Football Manager 2017, presented in a 10% frequency distribution plot

Figure 4.6 presents the monthly interval between the two games, and the average between them is 0,227934, which is very close to the average between the weekly correlation. Even with that fact, the pattern in this figure is a lot more different than the previous two.

Football Manager 2017 versus Final Fantasy XIV

This section contains the correlation values between Football Manager 2017 and Final Fantasy XIV, and they will be presented in daily, weekly and monthly intervals. These two video games are assumed to not be similar to each other, based on their difference in genres.

Football Manager 2017 versus Final Fantasy XIV - daily interval based correlation

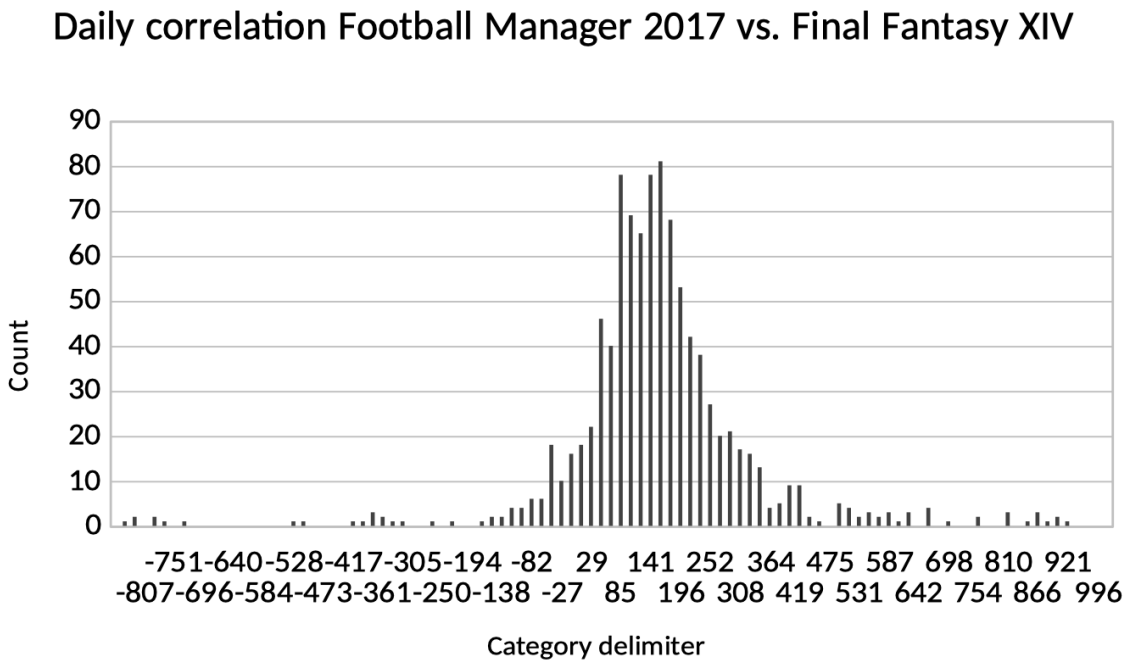

Figure 4.7: Daily correlation for Football Manager 2017 versus Final Fantasy XIV, presented in a 1% frequency distribution plot

Football Manager 2017 and Final Fantasy XIV are two games that were assumed to be different, based on prior knowledge. In Figure 4.7 above we see that the majority of the values are towards the centre of the figure, and the average in the daily intervals for these two games are 0,145664 which is quite close to 0. 0 would mean that there is no connection between these two patterns. Still, there is a long range in these numbers, as the lowest correlation is -0,862403 and the highest is 0,996105.

Football Manager 2017 versus Final Fantasy XIV - weekly interval based correlation

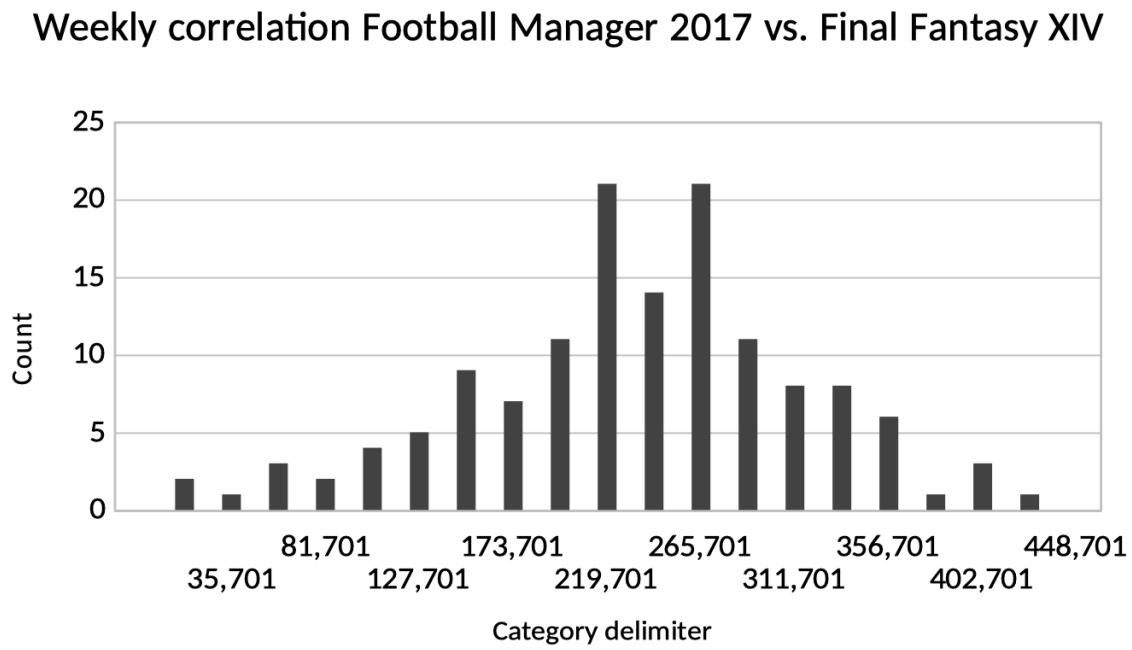

Figure 4.8: Weekly correlation for Football Manager 2017 versus Final Fantasy XIV, presented in a 5% frequency distribution plot

In Figure 4.8 we see a very similar pattern to Figure 4.7, as the curve starts on the lower side, are a lot higher in the middle before it goes back down. However, a big difference in the numbers in the weekly intervals is that the range is a bit smaller, with the highest being 0,449398 and the lowest is -0,009299. This is something that supports hypothesis H1, because the similarity changes as time changes. In addition, it also strengthens our assumption A1 because the pattern holds for a period of time in this case.

Football Manager 2017 versus Final Fantasy XIV - monthly interval based correlation

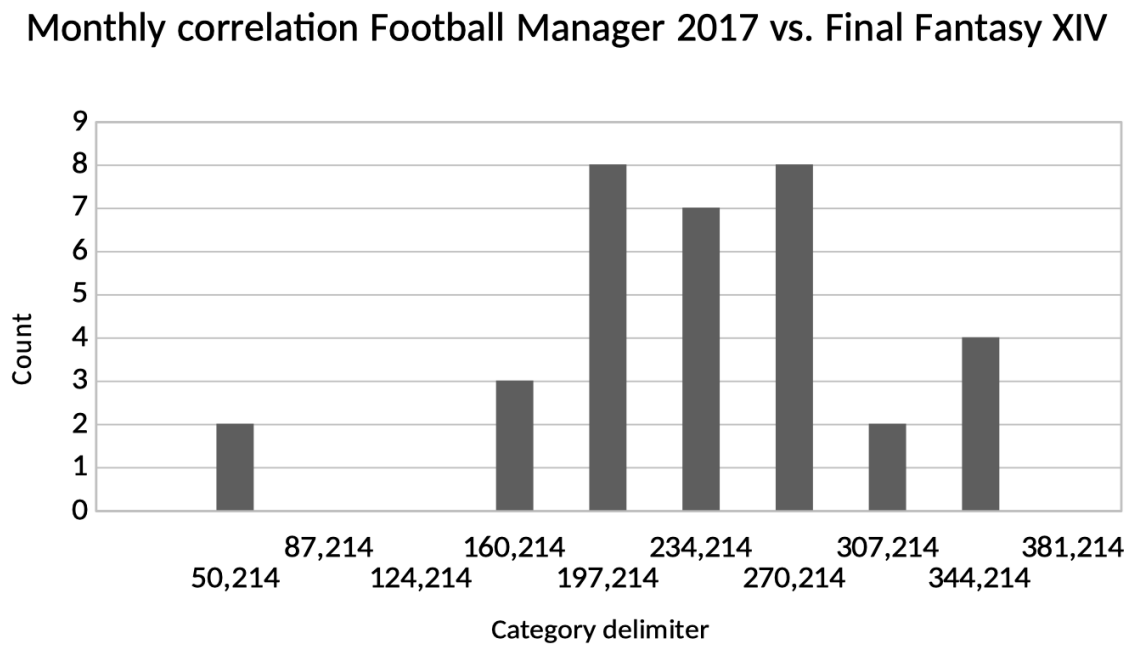

Figure 4.9: Monthly correlation for Football Manager 2017 versus Final Fantasy XIV, presented in a 10% frequency distribution plot

The pattern in Figure 4.9 is also a bit similar to Figure 4.7 and 4.8, because the highest points are in the middle of the figure, same as the other two. The average in the monthly intervals are 0,220557. So compared to the other intervals, this case is also quite close to 0 which means that there are a very small connection between the video games. The assumption between these two were as mentioned, that they would not be similar to each other as well.

4.2.4 Selected games: Total War: Warhammer II, Sid Meier's Civilization IV and Payday 2

In this group, the assumption is that Total War: Warhammer II and Sid Meier's Civilization IV are similar games, as they revolve around strategy, where the former is real time strategy and the latter is turn based strategy. Payday 2 is an action role playing game, and possibly more intense to play. As with the first group of video games, these are also a part of a series of video games, and all of them have at least one predecessor.

The correlation on calculated on the whole profile of these three games are as follows:

- The correlation coefficient between Total War: Warhammer II and Sid Meier's Civilization IV is 0,647372
- The correlation coefficient between Total War: Warhammer II and Payday 2 is 0,268359
- The correlation coefficient between Sid Meier's Civilization IV and Payday 2 is 0,265537

As initially stated, Payday 2 were assumed to be different from Total War: Warhammer II and Sid Meier's Civilization IV, based mostly on the difference in intensity of the video games and subgenre. As these video games are all action games, it is not surprising that all the correlation values are positive, even though Payday 2 correlates lower with the other two.

Total War: Warhammer II versus Sid Meier's Civilization IV

In this section, the correlation values between Total War: Warhammer II and Sid Meier's Civilization IV will be presented. The presumption is that these games are similar, as they are both strategy games which often are played for a long time. Firstly, a presentation of all the values will be presented, and then moving forward to daily, weekly and monthly intervals.

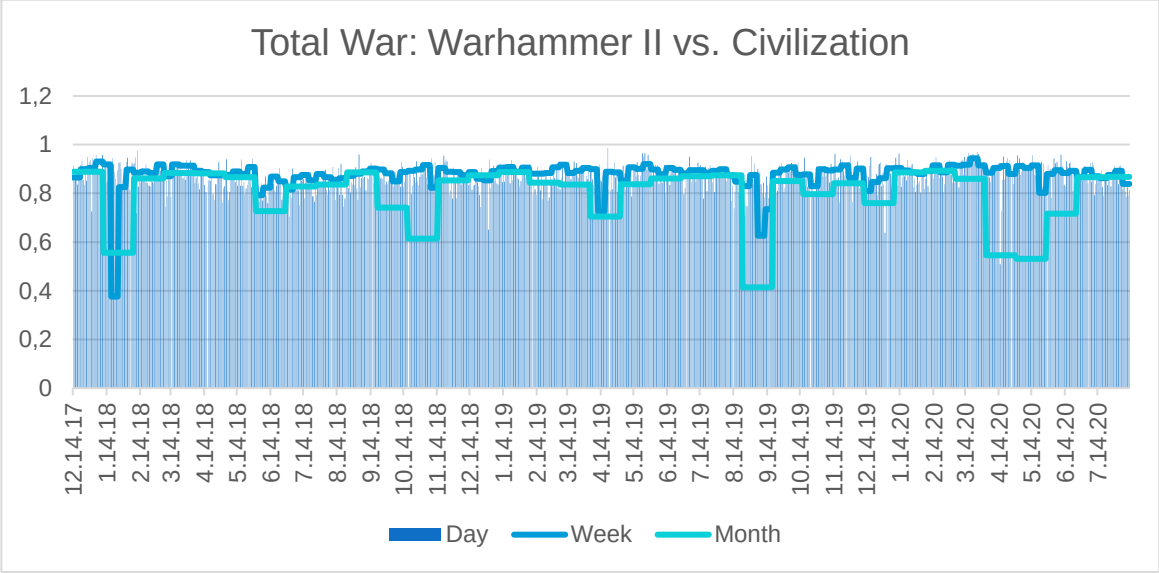

Figure 4.10: Daily, weekly and monthly correlation values for Total War: Warhammer II and Sid Meier's Civilization IV presented for de entire period of the dataset

Figure 4.10 shows a diagram representing all the intervals for Total War: Warhammer II and Sid Meier's Civilization IV. The daily intervals are shown as a vertical lines, and all days have a high positive correlation. In the horizonal lines weekly and monthly correlation is represented, in order to show the difference within the time intervals. As all days in this case have a high positive correlation value, it is natural that both weekly and monthly are high as well. These two intervals contribute to providing more information about the correlation values, as there are more variation in this, for instance, as shown in January 2018 and September 2019.

Total War: Warhammer II versus Sid Meier's Civilization IV - daily interval based correlation

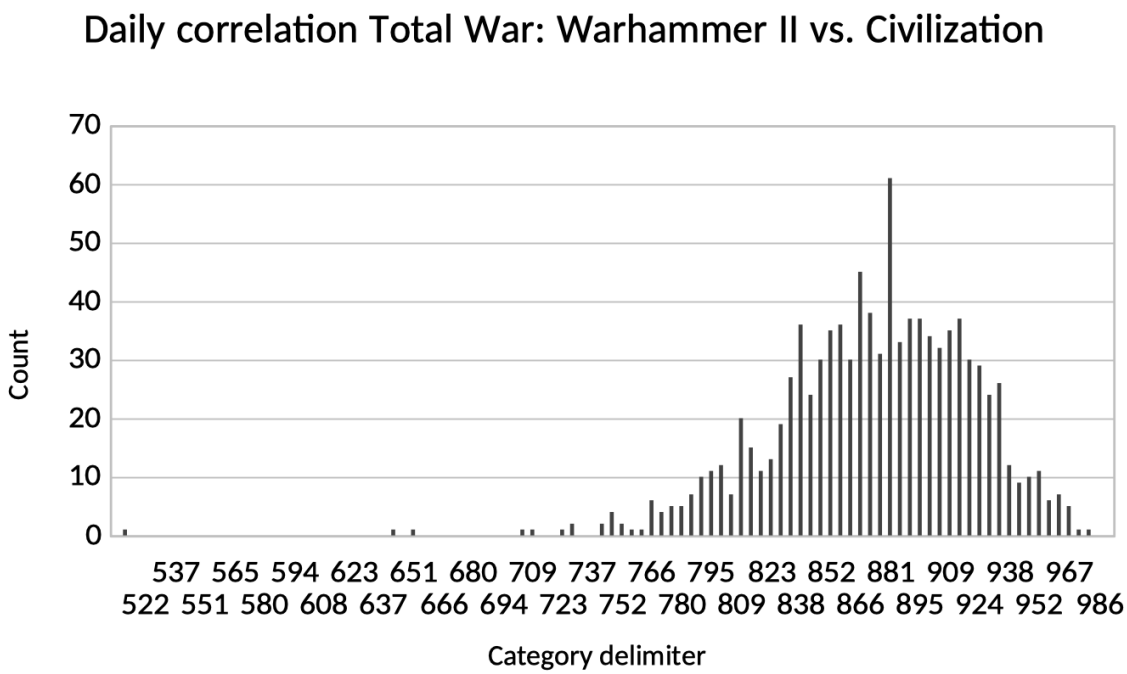

Figure 4.11: Daily correlation for Total War: Warhammer II versus Sid Meier's Civilization IV presented in a 1% frequency distribution plot

Figure 4.11 shows the daily correlation intervals for Total War: Warhammer II and Sid Meier's Civilization IV, these are two games that are assumed to be similar based on initial knowledge about them. The lowest correlation value between the two are 0,508589 and the highest is 0,986131. As stated earlier, these two games were assumed to be similar, and these results strengthens the assumption that similar video games behave similarly.

Total War: Warhammer II versus Sid Meier's Civilization IV - weekly interval based correlation

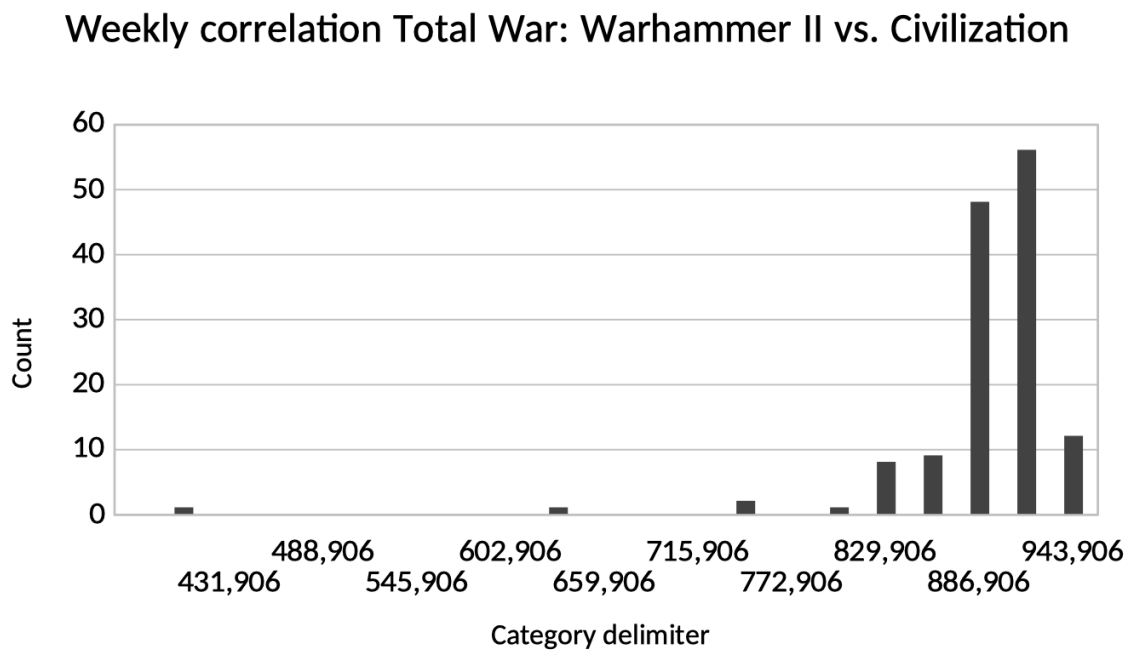

Figure 4.12: Weekly correlation for Total War: Warhammer II versus Sid Meier's Civilization IV presented in a 5% frequency distribution plot

As with Figure 4.11, the weekly intervals between these two games are also quite high, with the lowest correlation coefficient over the course of 973 days being 0,375906, and the highest being 0,943922. The average in this case is 0,876789 which is also quite high. This figure has a pattern that is a bit different from the pattern in Figure 4.10, but it still tells us that the correlation between the two games are positive and on the higher side in the correlation scale that is between 1 and -1.

Total War: Warhammer II versus Sid Meier's Civilization IV - monthly interval based correlation

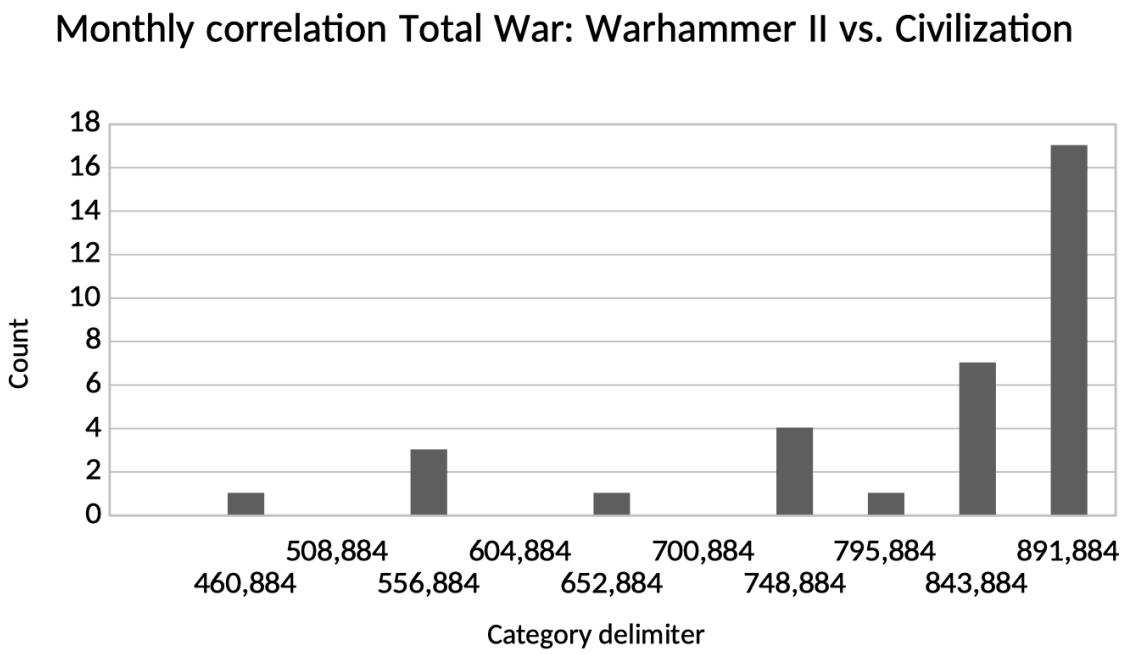

Figure 4.13: Monthly correlation for Total War: Warhammer II versus Sid Meier's Civilization IV presented in a 10% frequency distribution plot

Figure 4.13 presents the monthly intervals, for the same games as in Figure 4.11 and 4.12. Similarly to the weekly intervals, the highest value is 0,892586 and the lowest is 0,413884, and the average is 0,795418. This shows that correlation between the two games changes slightly when the interval is changed, in this particular case the change in these two is present, but it is not significant.

Total War: Warhammer II versus Payday 2

This section presents the correlation values between Total War: Warhammer II and Payday 2, these games were initially assumed to be different from each other. Although they have same genre which is action, there is a difference in subgenre and based on prior knowledge the reasoning is that Payday 2 might be a more intense video game.

Total War: Warhammer II versus Payday 2 - daily interval based correlation

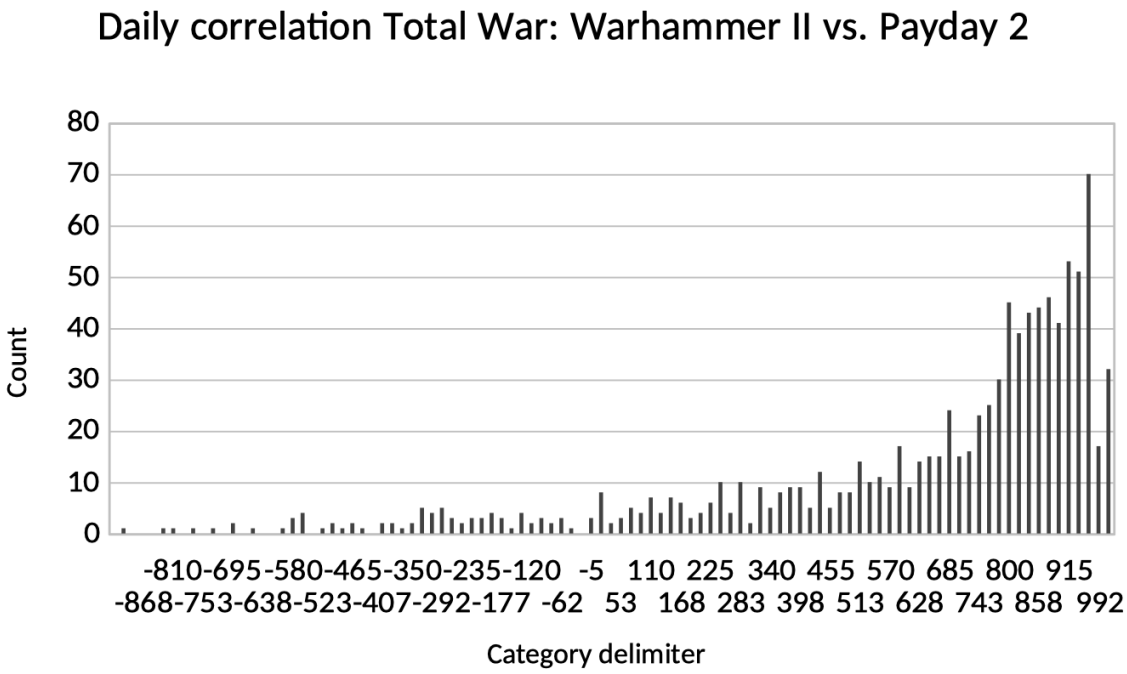

Figure 4.14: Daily correlation for Total War: Warhammer II versus Payday 2 presented in a 1% frequency distribution plot

Shown in Figure 4.14 are the correlations values for Payday 2 and Total War: Warhammer II, presented in a frequency distribution plot. These two games were initially assumed to not be similar to each other, but as the figure shows, the majority of the correlation coefficients are on the higher side. The lowest correlation value within this interval is -0,924574, the highest is 0,992409 and the average is 0,622932.

Total War: Warhammer II versus Payday 2 - weekly interval based correlation

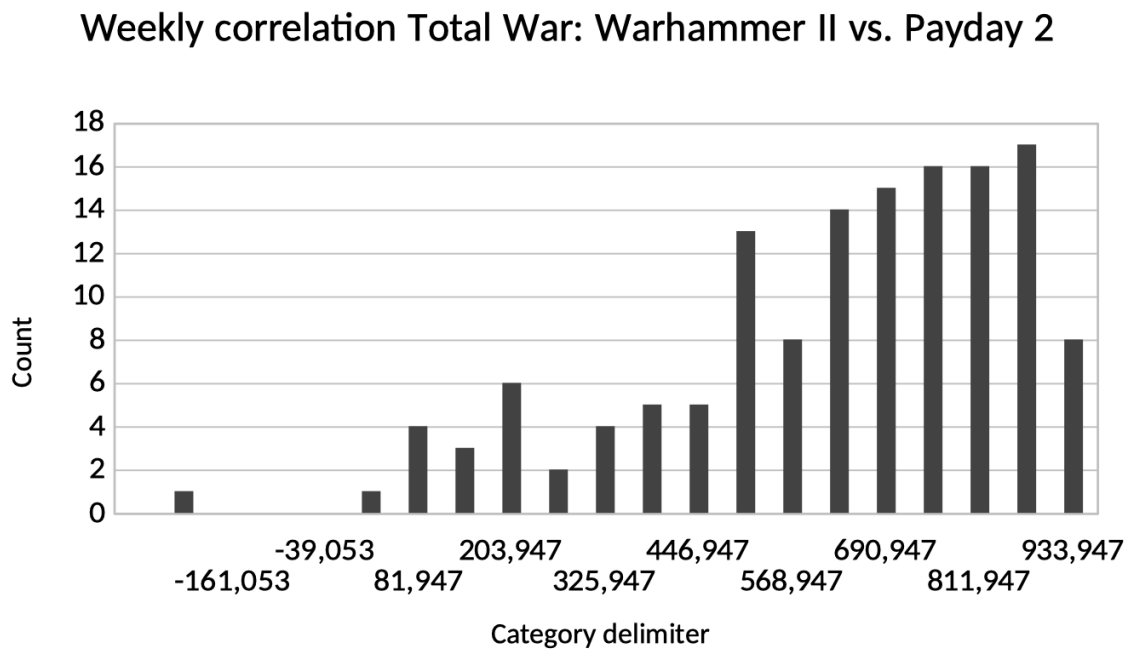

Figure 4.15: Weekly correlation for Total War: Warhammer II versus Payday 2 presented in a 5% frequency distribution plot

The weekly intervals for Total War: Warhammer II vs. Payday 2, as shown in Figure 4.15 above, changes slightly from the daily interval as shown in Figure 4.14. The distribution of the correlation coefficients moves a lot more to the left, which means that the correlation coefficients are a bit lower than they were for the daily interval, but they are overall still of quite high positive correlation.

Total War: Warhammer II versus Payday 2 - monthly interval based correlation

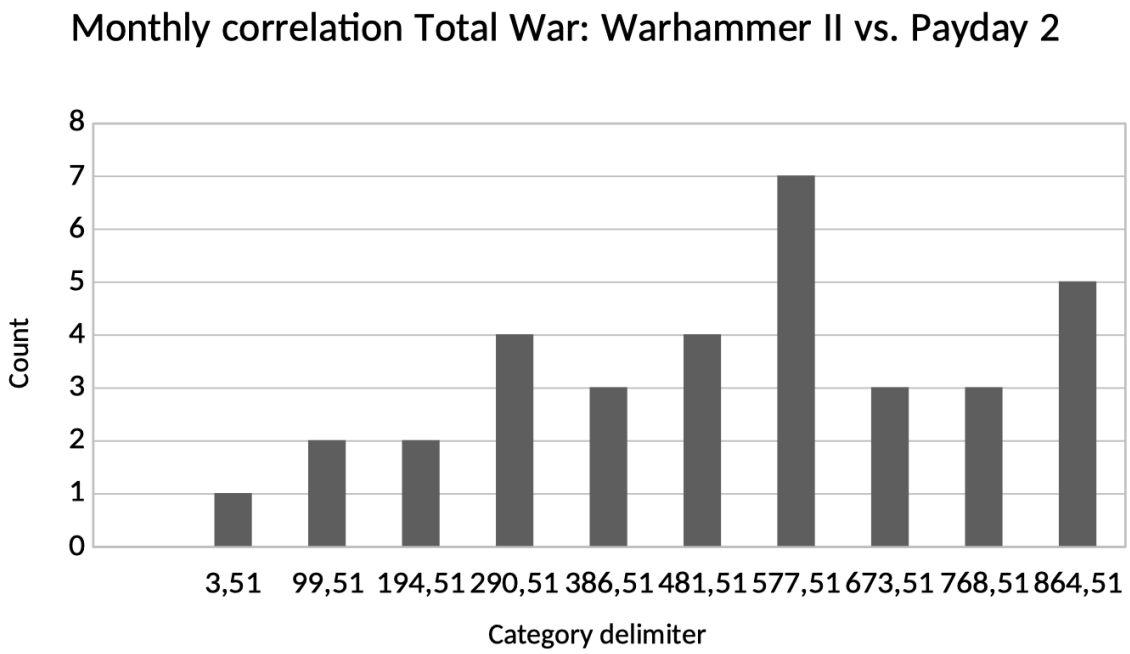

Figure 4.16: Monthly correlation for Total War: Warhammer II versus Payday 2 presented in a 10% frequency distribution plot

There is an even more visible change in the pattern in Figure 4.16, compared to both Figure 4.14 and 4.15. The values continue to shift towards left, and in this case there is one big top around 577,51, which originally is 0,57751 as the values have been normalized into hundreds. The split into intervals for these two video games is a good example where we see that the similarity changes with a change in intervals.

Sid Meier's Civilization IV versus Payday 2

This section will present the correlation values between Sid Meier's Civilization IV and Payday 2. First all the values will be presented in a figure, before moving on to daily, weekly and monthly interval splits.

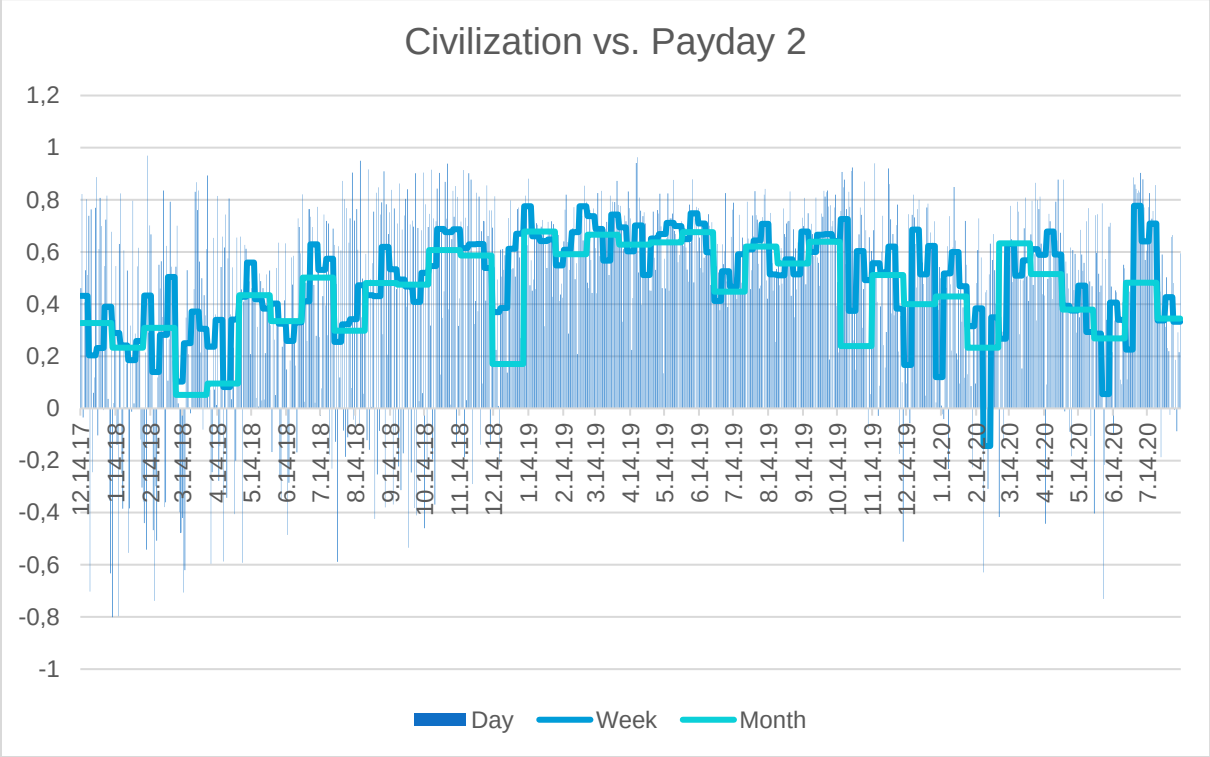

Figure 4.17: Daily, weekly and monthly correlation values for Sid Meier's Civilization IV and Payday 2 presented for de entire period of the dataset

As shown in Figure 4.10, Figure 4.17 also presents the correlation values for days, weeks and months. The daily values are presented in vertical lines, and weekly and monthly are presented in horizontal line. The two video games in this case are Sid Meier's Civilization IV and Payday 2, which are two video games that were assumed to be different from each other. The reason for this is that Sid Meier's Civilization IV is a turn based strategy game, which often are played for a long period of time, on the other hand, Payday 2 is a cooperative action video game which can be seen as quite intense with a higher pace. Still, these video games appear as similar through high positive correlation values at certain points.

Sid Meier's Civilization IV versus Payday 2 - daily interval based correlation

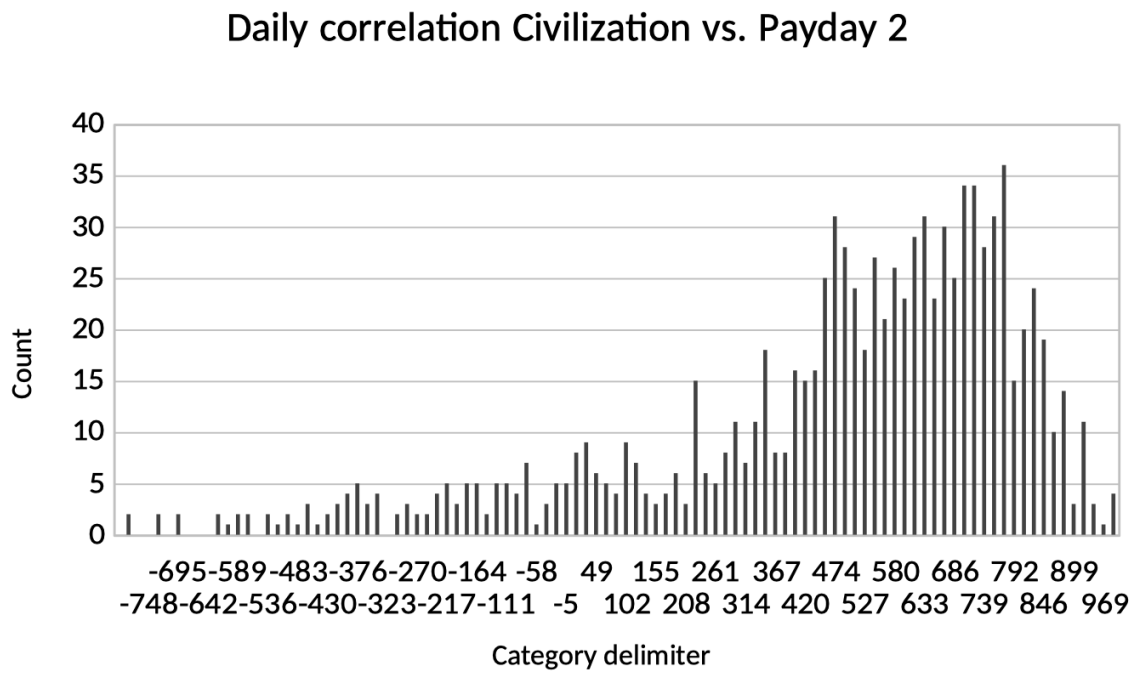

Figure 4.18: Daily correlation for Sid Meier's Civilization IV versus Payday 2 presented in a 1% frequency distribution plot

Figure 4.18 presents the daily interval correlations for Sid Meier's Civilization IV, which are two games that initially were assumed as not similar. As stated earlier, these numbers have been normalized into hundreds, in order to present the correlation coefficients in these numbers. The reason for the presumption that these video games are not similar, is that Civilization is a strategy game, while Payday 2 is a first person shooter that is more intense. There is a possibility that Civilization is a game that is played for a longer period of time during a session, while Payday 2 might be played for a shorter period because of the intensity.

The results in this case is surprising, because the average between the two games are 0,462325. The highest correlation value in this interval is 0,969984, while the lowest is -0,801016, this tells us that there obviously is a long span between these correlation values. Still, the average is as high as 0,462325, which again means that there are mostly correlations that are on the higher side between these two video games.

Sid Meier's Civilization IV versus Payday 2 - weekly interval based correlation

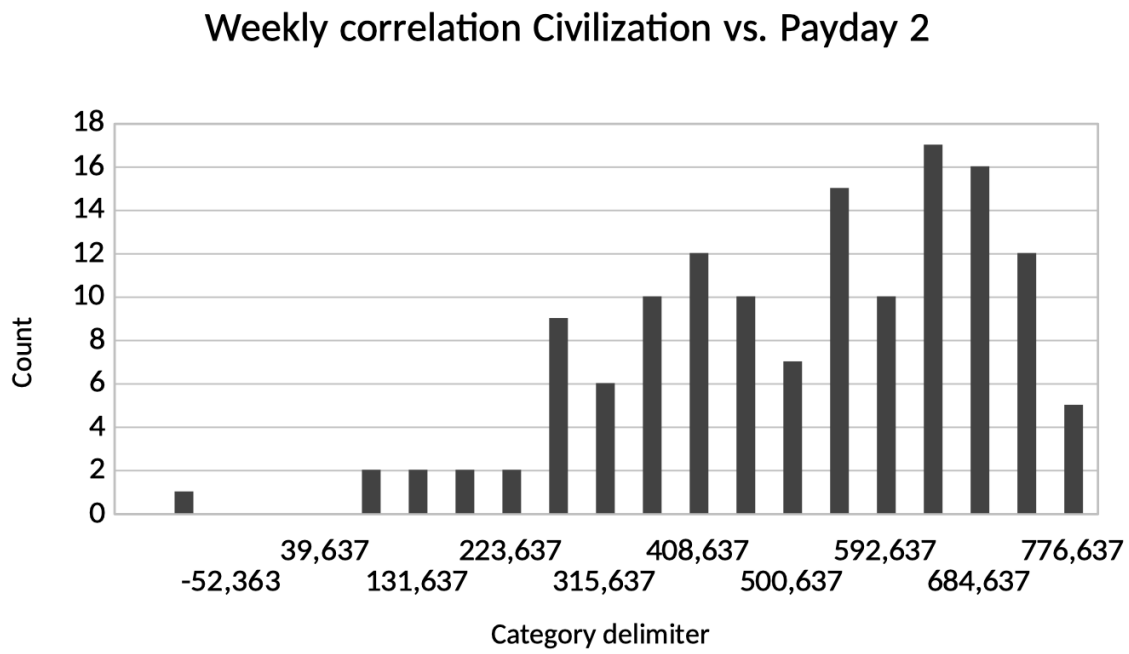

Figure 4.19: Weekly correlation for Sid Meier's Civilization IV versus Payday 2 presented in a 5% frequency distribution plot

Figure 4.19 shows the weekly intervals for Civilization and Payday 2, and this figure follows the same pattern as Figure 4.18. The average between these two games are 0,488787, which is a bit higher than the average for the daily intervals. But still, as seen from both these figures, they have a very similar pattern, even though the interval has changes. These changes mainly involve that the highest correlations within this interval is lower than in the daily interval, but as stated the pattern is still quite similar.

Sid Meier's Civilization IV versus Payday 2 - monthly interval based correlation

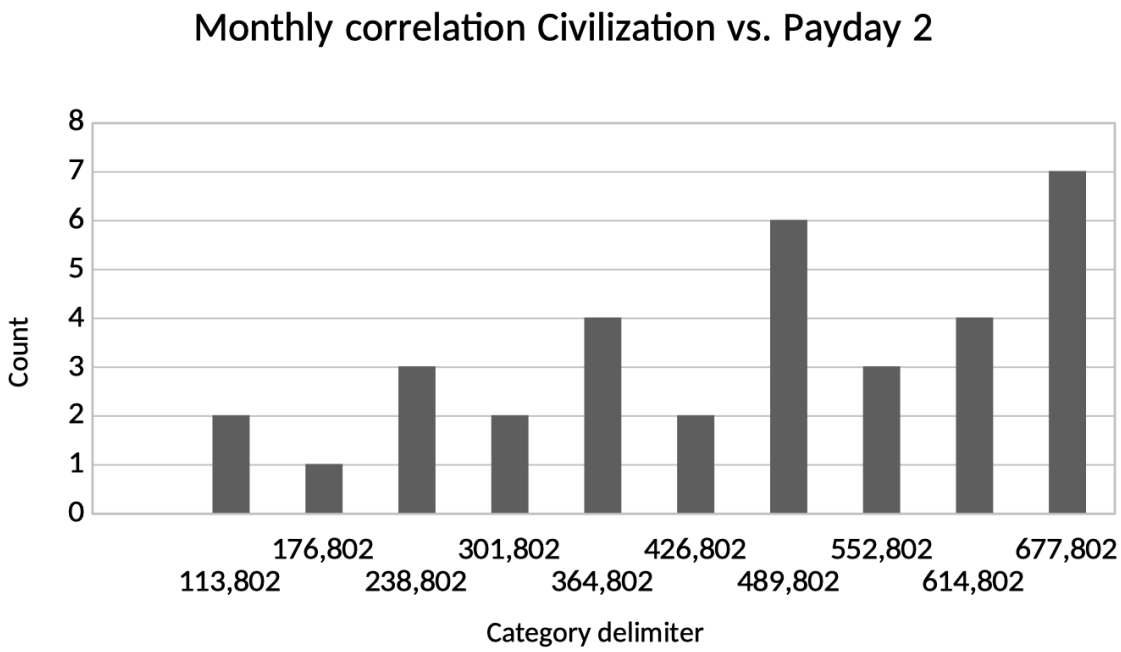

Figure 4.20: Monthly correlation for Sid Meier's Civilization IV versus Payday 2 presented in a 10% frequency distribution plot

Figure 4.20 has a slightly different pattern than the two former revolving around these two games, and the probable explanation for this is that the interval has changed and is a longer interval than the other two presented. An interesting fact though, is that the average in this case is 0,442294 which is close to the average for both the daily and weekly intervals. The reason for this may just be that the highest and lowest correlation value in the monthly intervals does not have that big of a range.

Complete dataset correlation – PlayercountHistoryPart1

After conducting complete correlation the result is approximately half a million correlation coefficients, and below the 25 highest correlating video games will be presented, before moving on to the 25 lowest correlating video games within playercounthistorypart1.

Highest 25 correlating video games

	Video game 1	Video game 2	Correlation
1	Football Manager Touch 2018	Football Manager 2018	0.993234
2	Football Manager Touch 2017	Football Manager 2016	0.986254
3	Football Manager 2012	Football Manager 2013	0.986194
4	Counter-Strike: Source	Counter-Strike	0.986194
5	Nobunaga's Ambition: Souzou with Power Up Kit	NOBUNAGA'S AMBITION: Souzou Sengoku Risshiden	0.981702
6	Football Manager Touch 2017	Football Manager 2017	0.980575
7	Empire: Total War	Napoleon: Total War	0.979576
8	Medieval II: Total War	Rome: Total War	0.979522
9	Star Wars: Knights of the Old Republic II: The Sith Lords	Star Wars: Knights of the Old Republic	0.978976
10	Pro Evolution Soccer 2017	Football Manager 2017	0.978137
11	Pro Evolution Soccer 2018	Pro Evolution Soccer 2018 Lite	0.976693
12	Football Manager 2014	Football Manager 2015	0.974792
13	Football Manager 2015	Football Manager 2016	0.97439
14	Football Manager 2014	Football Manager 2013	0.974046
15	Portal 2	Portal	0.971686
16	Football Manager 2017	Football Manager 2016	0.971358
17	Sid Meier's Civilization IV: Beyond the Sword	Sid Meier's Civilization IV	0.970574
18	Age of Empires III: Complete Collection	Age of Mythology: Extended Edition	0.969979
19	Midas Gold Plus	Transport Defender	0.969677
20	Sid Meier's Civilization IV: Beyond the Sword	Sid Meier’s Civilization III: Complete	0.966034
21	Age of Mythology: Extended Edition	Rise of Nations: Extended Edition	0.965394
22	NBA 2K17	NBA 2K16	0.963011
23	Saints Row IV	Saints Row: The Third	0.962254
24	The Witcher 2: Assassins of Kings Enhanced Edition	The Witcher: Enhanced Edition	0.962061
25	Empire: Total War	Rome: Total War	0.961752

Table 4.4: The 25 highest correlating video games in playercounthistorypart1

Presented in Table 4.4 are the 25 highest correlations within playercounthistorypart1 in the dataset, it is apparent from this data that there are a lot of different franchises that appear throughout these results, especially the Football Manager franchise. In addition to that, we see that the majority of these games are predecessors and successors within the same series. At line 17 in Table 4.4, we see that video game 1 in that row is a DLC to video game 2 in the same row. This is interesting because a DLC is an expansion pack for a video game, which means that there is additional content added to an already existing video game, and there is a requirement to own the original game in order to buy and/or download the DLC. However, in this case the explanation is possibly that many people simply does not buy the DLC, and continue to play the original video game. Therefore, the players that have bought the DLC will be registered as playing that as a video game in Steam, while the players that do not have the DLC will be registered as playing the original video game.

Lowest 25 correlating video games

	Video game 1	Video game 2	Correlation
1	Ragnarok Journey	DayZ	-0.713437
2	Ragnarok Journey	The Isle	-0.707316
3	DayZ	Granado Espada	-0.67917
4	DayZ	Magic Duels	-0.669146
5	The Isle	Granado Espada	-0.668817
6	The Isle	Magic Duels	-0.668304
7	MOBIUS FINAL FANTASY	Wallpaper Engine	-0.664077
8	Ragnarok Journey	3D Mark Demo	-0.658485
9	Wallpaper Engine	Realm Grinder	-0.657084
10	The Isle	Zombidle: REMONSTERED	-0.650256
11	Granado Espada	Wallpaper Engine	-0.648474
12	The Isle	Midas Gold Plus	-0.646879
13	3D Mark Demo	Winning Putt: Golf Online	-0.643136
14	MOBIUS FINAL FANTASY	Discord Bot Maker	-0.642117
15	3D Mark Demo	Franchise Hockey Manager 4	-0.638633
16	DayZ	Midas Gold Plus	-0.63447
17	3D Mark Demo	Magic Duels	-0.62683
18	The Isle	Winning Putt: Golf Online	-0.625587
19	Magic Duels	Wallpaper Engine	-0.6244
20	Fallen Earth	Wallpaper Engine	-0.623337
21	The Isle	MOBIUS FINAL FANTASY	-0.622954
22	MOBIUS FINAL FANTASY	Blender	-0.621779
23	The Isle	Transport Defender	-0.621386
24	NBA 2K18	3D Mark Demo	-0.621329
25	Granado Espada	The Drone Racing League Simulator	-0.620774

Table 4.5: The 25 lowest correlating video games in playercounthistorypart1

Table 4.5 presents the 25 video games with the lowest negative correlation in the first part of the dataset, and can be seen as video games that are different to one another. Whether or not they actually are different from each other, is not something that can be established until they are investigated even further, but still contributes to a better overview when it comes to similarity.

In this table there are several different applications that are tools, rather than video games, for instance 3D Mark Demo, Wallpaper Engine and Discord Bot Maker, and therefore it is not surprising that they have a high negative correlation coefficient when compared to applications that actually are video games.

Complete dataset correlation – PlayercountHistoryPart2

Similarly to *playercounthistorypart1,* the same complete correlation has been conducted on *playercounthistorypart2,* and below the 25 highest correlating video games and the 25 lowest correlating video games in *playercounthistorypart2* will be presented.

Highest 25 correlating video games

	Video game 1	Video game 2	Correlation
1	The Jackbox Party Pack	The Jackbox Party Pack 2	0.987755
2	The Jackbox Party Pack 2	Quiplash	0.985592
3	Half-Life 2: Episode One	Half-Life 2: Episode Two	0.980866
4	The Jackbox Party Pack	Quiplash	0.97405
5	The Jackbox Party Pack 2	The Jackbox Party Pack 4	0.971385
6	Gothic II: Gold Edition	Gothic 3	0.966935
7	Ticket to Ride	Catan Universe	0.960674
8	Football Manager 2010	Football Manager 2011	0.959535
9	Football Manager 2010	Football Manager 2009	0.959295
10	The Jackbox Party Pack	The Jackbox Party Pack 4	0.95826
11	Risen 2 – Dark Waters	Risen 3 – Titan Lords	0.958201
12	The Walking Dead: Season 2	The Walking Dead	0.95632
13	Space Pirate Trainer	The Lab	0.956001
14	The Walking Dead: A New Frontier	The Walking Dead: Season Two	0.955866
15	Wolfenstein: The Old Blood German Edition	Wolfenstein: The New Order German Edition	0.954752
16	S.T.A.L.K.E.R.: Call of Pripyat	Dishonored	0.953204
17	The Walking Dead: A New Frontier	The Walking Dead	0.952772
18	MechWarrior Online	Football Manager 2009	0.948344
19	Quiplash	The Jackbox Party Pack 4	0.947287
20	Might & Magic Heroes VII Trial by Fire	Might & Magic Heroes VII	0.9462
21	Call of Duty: Advanced Warfare - Multiplayer	Call of Duty: Ghosts - Multiplayer	0.945788
22	Danganronpa V3: Killing Harmony	Danganronpa 2: Goodbye Despair	0.944307
23	Darksiders Warmastered Edition	Darksiders II Deathinitive Edition	0.944284
24	Pharaoh + Cleopatra	Caesar 3	0.943994
25	Guild Quest	Ragnarok Clicker	0.943351

Table 4.6: The 25 highest correlating video games in playercounthistorypart2

In Table 4.6 above, we see the 25 highest correlating video games in part two of the dataset, which is 1001-2000 amongst the 2000 top played applications in Steam at

the point the data were retrieved. As with Table 4.4, that presented a lot of games within the franchise concept, the same concept appears within these video games. Amongst the top 5 in this table, we mainly see different versions of The Jackbox Party Pack, which is a social game consisting of different video games that people often play together as a social activity. Quiplash is in a way similar to a DLC, equivalently to the case with Sid Meier's Civilization IV and Sid Meier's Civilization IV: Beyond The Sword, as found in Table 4.4 The difference here is that Quiplash is a mini game that is a part of The Jackbox Party Pack, but can be bought and played on its own, as a contrast to Beyond The Sword.

The most surprising result in Table 4.6 is on line 18, where we find MechWarrior Online and Football Manager 2009. These are two seemingly different video games in terms of theme, as Football Manager is a sport video game and MechWarrior Online is a video game based on controlling combat vehicles [65], but they are both simulation video games.

Lowest 25 correlating video games

	Video game 1	Video game 2	Correlation
1	Duelyst	ShareX	-0.748412
2	Ragnarok Clicker	Aseprite	-0.742332
3	Ragnarok Clicker	ShareX	-0.741828
4	Duelyst	Aseprite	-0.73134
5	ShareX	Atlas Reactor	-0.694433
6	Guild Quest	ShareX	-0.689006
7	Gloria Victis	ClickRaid	-0.686639
8	WildStar	Kenshi	-0.67667
9	Guild Quest	Aseprite	-0.676254
10	Forge of Gods (RPG)	Kenshi	-0.675245
11	ClickRaid	Kenshi	-0.668494
12	Guild Quest	Kenshi	-0.664953
13	WildStar	Gloria Victis	-0.644
14	Aseprite	Atlas Reactor	-0.663025
15	sZone Online	Atlas Reactor	-0.661635
16	GameMaker Studio 2 Desktop	Ragnarok Clicker	-0.660551
17	WWE 2K18	Kenshi	-0.657449
18	Art of War: Red Tides	ShareX	-0.655624
19	Forge of Gods (RPG)	ShareX	-0.654752
20	ShareX	Blacklight: Retribution	-0.654463
21	Art of War: Red Tides	Aseprite	-0.65058
22	Puzzle Pirates: Dark Seas	ShareX	-0.649386
23	Ragnarok Clicker	Zuma Deluxe	-0.646395
24	Duelyst	Zuma Deluxe	-0.64626
25	Gloria Victis	WWE 2K18	-0.645316

Table 4.7: The 25 lowest correlating video games in playercounthistorypart2

Table 4.7 presents the 25 lowest correlating video games in part2 of the dataset, which similarly to Table 4.5, includes several applications that are tools rather than video games. The same statements as were made with Table 4.5 is also true here, initially the assumption is that these video games/applications are different from one another, but further research is necessary in order to determine that for certain.

4.3 Chapter summary: revisiting the theoretical model

In this section we will revisit the assumptions and hypotheses that were developed as a part of our theoretical model, to create some foundation for further investigation. The assumptions will be presented first, with information in regards to what we have found or not found, before moving on to the hypotheses.

4.3.1 Revisiting assumptions

> **A1:** *All games **g** have a periodic pattern **p** which holds for a period **t***

This assumption has not been verified, the reason for that is that none of the video games have been compared to themselves, they have only been compared to others. The video games that have been compared to each other are periodic, because video games in general resemble each other, especially if they are divided into days. This could have been done by comparing a video game against itself over several days, but would not have the same pattern for the whole period, as this is something that changes over time.

> **A2:** *Every pattern **p** has a similarity **s** with every other pattern which can be established numerically*

Assumption A2 has been used throughout the entire project, but not for any other pattern than within the intervals that are days, weeks and months. These patterns have been established with correlation.

> **A3:** *For patterns **p** and **p'**, the similarity function **s**(p,p') would return a numerical value describing their similarity*

This can be seen in context with A2, because both assumptions are based on numerical values, and this has been done in this project with correlation. However, A3 does not deal with how to handle missing values in the dataset.

> **A4:** *Based on this similarity **s**, groups of games can be established*

We believe in this assumption, even though it has not been tested or verified, which is something that can be beneficial to do in future work related to this research. Principle component analysis and other component analysis forms can still be conducted in the future. In this project the concept of similarity was found to be more challenging than initially anticipated, which again lead to more work and time spent on this concept.

4.3.2 Revisiting the hypotheses

H1: *The similarity **s** of two patterns change with different values of **t***

This hypothesis has been verified, and it is quite clear from the research that has been conducted, especially within the narrow approach and the 12 selected games where different intervals were tested.

H2: *Groups of similar games change with different values of **t***

This can partially be confirmed, although it is not verified in this project, because we have not conducted any clustering algorithms. Still, some groups have been identified, which are qualitatively similar. For instance, franchises like Football Manager and other applications that are more tools than they are video games, like ShareX, Aseprite and Wallpaper Engine. These are groups we can read from our results, and they are not the results of a formula or an algorithm, and therefore we do not know if this is something that changes with time.

One thing we can read from the results that are related to time, is that the franchise video games, especially Football Manager, changes over time. This is something that we have learned because from the results in Table 4.4, the video games in the franchise only have a high correlation value with the version before or the version after, we have not found that a version that is a few years old correlate highly with the newest version.

H3: *A high similarity over a long **t** does not guarantee a high similarity for a smaller **t'** within the original **t***

In this hypothesis, we have observed the opposite, where the correlation value is higher in a shorter interval of time, than what the correlation is when looking at the whole profile of the video games. This is a hypothesis that cannot be verified, only falsified, and in order to do that we would have to investigate intervals in all video games in the dataset, to get a specific result.

As mentioned earlier, correlation was calculated in the intervals of 6 months and a year as well as daily, weekly and monthly intervals. These calculations follow the same trend for all the video games, where the correlation gets lower and lower as the interval becomes larger. This is something that all video games where this calculation has been done, have in common. These video games are NBA 2k18, Football Manager 2017, Final Fantasy XIV, Total War: Warhammer II and Sid Meier's Civilization IV, which are the video games found in the two groups of selected games that have been explained more in depth. This also supports assumption A1, that all games have a pattern that which for a period of time.

With the foundation that is provided by these assumptions and hypotheses, it contributes to exploring some important things in regards to this research, as well as it works like a guideline. Although all of these have not be answered to their full extent, this is something that could be done in context with future work.

Chapter 5 Discussion

5.1 Reflection on approaching the dataset

The dataset in this project is extensive, and it includes information about player count for two thousand video games and a timespan of approximately two and a half years. The first category which is playercounthistorypart1, includes data from every fifth minute in the time span. The other category, which is playercounthistorypart2, includes data for every hour, as these games are played less than the video games in the former folder. Nonetheless, this research revolves around an enormous amount of data and the result is an exploratory approach to the project.

Comparison of very single file to one another, would have been the alternative approach, and a very extensive method as such. The consequence of choosing a more specific approach is more narrow-mindedness within the project, and whilst that is recognized, it might have been very overwhelming. Therefore, the initial thoughts were to select a set of games that were assumed to be similar, or contrasting. This method was chosen in order to get an overview and a better insight in regards to assumptions and prior knowledge of video games in general.

As mentioned, this dataset is extensive, and there are 2000 video games in total and as many files involved in the calculation of correlation coefficients, it has been difficult to structure a way of attack when it comes to running through all the data and the organization of the process. The 2000 is as mentioned split into two different parts, so when it comes to the complete correlation it concerns 1000 files in two different processes. A consequence of this was that the process was somewhat disorganized in the beginning, and it became important to contemplate the best way to move forward for handling this amount of data. This involved starting and stopping this particular process a few times, but as a result there is now a better control of the data and the processes revolving around retrieving the results. The knowledge gained in regards to this is important for future work, because it provides more knowledge and preparation for conducting further research on this subject. It can be quite difficult to

navigate through a big dataset as the one used in this project, so it is important to prepare better for handling this amount of data so that the results are not too overwhelming. This includes that it the calculation of each group of 1000 games could have been prepared in a better way, possibly by focusing on a selection of 20 games to begin with in order to make sure the script would work as expected before running through all 1000 video games. Still, the results needed have been calculated and therefore the purpose has been fulfilled, but a different way of approach into this may have led to better structure and overview.

5.2 Possibilities for future research on the subject

The central thought in this case is the subject of similarity with using correlation to give us an understanding as to if an amount of players can tell us something about how video games can be classified as similar. Future work on the subject can involve clustering algorithms and possibly help us gain more knowledge on the subject, from a different point of view. The resources in this particular project revolve around one particular platform providing video games, *Steam*. Even though *Steam* provides a collection of video games in a variety of genres, where some of them are certainly some of the most played and well known games in general. There are also other platforms that provide video games that are popular to the same extent, perhaps for a different audience. Therefore, it will also be interesting to look into the possibility of retrieving this kind of data from other platforms as well in order to gain more knowledge from an even broader perspective.

As the video game industry is continuously undergoing change and development, both in terms of new games, consoles, devices, developers and the community itself. The video game community is increasing all the time and there is more focus on the positive sides of this industry, as a contrast to the negative sides that have been highlighted throughout the years. Meaning that there have often been discussions on the negative sides of video games, for instance when it comes to violent aspects and the influence that might have [12]. On the other hand, there have recently been more positive views in regards to both social aspects and more focus on video games that can be used for learning, examples of this are strategic games and language, for countries where English is not the first language [66]. Research into the field can be

beneficial for developing even more knowledge of this, both with determining similarity and possibly future work that can contribute to gaining even more information about the subject.

These factors does not necessarily mean that there is a straightforward way to use the results and conclusions in this particular matter, because there can be several factors affecting this.

A new problem statement for future research is:

- What are the main factors that makes video games within the same genre, different?

This problem statement represents the same as in this thesis, but assumes the opposite. In this type of approach, the idea is to work the other way around, but conducting examinations like the narrow approach based on findings in the broad approach. This can contribute to finding different aspects, and history within this video games. Essentially, this is about mapping instead of mainly working with numbers.

5.3 The different approaches in this research

There have been two different processes in regards to approaching the dataset, and for deciding the best approach there has been some difficulties as there are both positive and negative sides to the approaches, as discussed earlier. The two main ideas in this case have been a broad approach, or a more narrowed approach. During the course of this research, both methods have been applied.

5.3.1 Broad approach to the dataset

By choosing a broader approach, the result will be a lot of data that can be processed, but the question in regards to this is what do to with these results when we have them. This type of broad approach involves calculating correlation on all files against each other, where the end results should be 499 500 correlation

coefficients, as the script for this would calculate the first game against 999 others, and then the next one against 998, and so on.

The formula for this is:

$$Sn = \left(\frac{n}{2}\right) * [2a + (n-1)d]$$

Where a = 1, d =1 and n = 999. By inserting these values into the formula, this is the result:

$$Sn = \left(\frac{999}{2}\right) * [2(1) + (999-1)(1)]$$

$$Sn = \left(\frac{999}{2}\right) * [2 + 998]$$

$$Sn = \left(\frac{999}{2}\right) * 1000$$

$$Sn = 499{,}500$$

This means that there is an overwhelming amount of data to sort through, it can be difficult to navigate through this amount of correlation coefficients. Something that possibly can be done, is a manual inspection of this data, but that again, rises a question of what we would actually gain from that. It is still interesting to see what kind of games that have high correlation, and which games that have low correlation. It is also important to consider that if several video games have 0 players or a low amount of players, the correlation would still be approximate to perfect correlation.

This method of approach has been quite time-consuming, both when it comes to the computer capacity, where it is clear that it would have been more efficient with better processors when running the script in order to perform this analysis. Another difficulty

when it comes to this method is that there was spent a lot of time trying to get it to run correctly. First of all, there were some struggles when it comes to the running of the script and getting it to write to the correct file, as well as not overwriting the previous information to this file. As this is a task that takes quite a lot of time, it is important to have the possibility to be able to stop the script and start it again without losing any important information. This was something that were not considered thoroughly enough before beginning the process, and therefore it resulted in some difficulties that could have been avoided to begin with.

In order to calculate all the files against one another, meaning the files in the first group of the dataset, which is PlayercountHistoryPart1 and PlayercountHistoryPart2, a script was created to be able to perform this operation. There are 1000 files, one for each game in each of these folders, where part1 contains data for every fifth minute, and part to contains data for every hour. Therefore, each file in part 1 has approximately 300 000 rows of data, and each file in part 2 has approximately 25 000 rows of data, but both produce the same amount of results as there is 1000 video games in each part. The result of this is that the first part takes a huge amount of time to process, and while the second part is smaller, there is still a lot of data to go through, so these two operations where quite long. The operation was performed on the first part of the dataset first, as it were expected to take quite a lot of time, and therefore it was at the time decided that it would be best to have those results first. In addition to this, it is somewhat more interesting to look at part 1 as it contains the 1000 most played video games on Steam.

Unfortunately there were some errors, miscalculations and other events that were not accounted for beforehand. For instance, there are some files that have missing data which leads to being able to process the calculations for the missing games, and a correlation value presenting as N/A. After further investigation the results with this presented value came as a result of the id *335360*, which is presented with the name SteamDB Unknown App 335360, and after checking this file there are no values in this particular file, as would be expected by the name. There has also been some incidents of correlation values that are higher than 1, but when running the games with this value again, they had an expected correlation value between 1 and -1. The reason for this is unknown, and there has not been done more investigation into this

at this moment. Therefore, the reason for the error in this occurrence is unknown. This lead to having to start and stop this particular script multiple times, for instance, to avoid the results being written over, as well as making sure it could run continuously and then be written to file with the information that was needed. This information being the id of the games that were calculated against each other and the correlation coefficient. In hindsight, it might have been a better approach to start with part2, as this script took a significantly shorter amount of time to complete. The initial thought was to start with part 2, but it was decided to start with part 1 because of the amount of data in those files. The reasoning behind this was that the video games in part 1 might be more interesting to look at, as they are the top applications, so we wanted to make sure that this process was completed. It would have been beneficial to conduct the testing and adjustments to the script when working with part 2 as it contains less data, but there are advantages to both approaches. There are two different purposes in regards to this method of approach, firstly when it comes to the distribution of results, and secondly, facilitation for further research that involves clustering algorithms, for instance principal component analysis, as mentioned earlier. The pathway to performing PCA is made through the dataset that has been developed in this research, as there are correlation values for all the video games that are included in the original dataset that includes playercount.

5.3.2 Narrow approach to the dataset

The other approach to the project was a more narrow-focused approach with a selection of video games that we assumed to be similar, and these games were selected only based on some prior knowledge about these video games. This approach contributed to a lot of testing in regards to the different scripts, and allowed us to gain more insight about the dataset moving forward. The first steps in this approach was to calculate correlation on the whole profile, the same way it has been done in the broad approach, with the exception that only two and two files were being processed at the time. These files were selected beforehand, because of the fact that we assumed some video games to either be similar or not.

Moving forward we wanted to look at how the data behaves when it is split into smaller intervals, over different periods of time. This is an approach that contributed

to gaining more knowledge about the data, as the intention was to begin with. One thing that is clear from this approach is that the data behaves differently, when it has been split into days, weeks and months. For instance, a video game might have a correlation around 0.5 when the whole profile is being processed, but our research shows that when it is split into months, the correlation might be higher in some instances. Furthermore, the correlation is usually even higher when it comes to weeks, and from that even higher when it is split into days. This also includes the other way around, for negative correlation. A combination of video games might have a correlation value closer to 0, for instance, -0.1, but when the data is split in to smaller groups, the negative correlation will appear as even closer to -1.

The positive side of choosing a narrow approach is, as mentioned that we gain more insight into some parts of the dataset itself, and this again helps us to consider what path we think will be beneficial moving forward. This also gives us an opportunity to further develop the scripts being used in this process, and refine this approach. There is a lot of room for learning here, and even though there is a lot to learn, the drawback is still that we learn a lot but about a little. The findings in regards to this is also something that cannot directly be transferred to the rest of the data, meaning that we cannot prove or guarantee generalizable results from this type of approach.

Another side of this process is that in some ways it can be a more rewarding way to work with the data. The meaning of this being more giving is that it provides a lot of knowledge and room for learning underway, as well as being able to test some theories we had before starting this research. This again, offers a lot of room for improvement, before moving forward to the next part of the project and a better understanding when it comes to the dataset and behaviour of the data in general. Another difficulty with a more narrowed approach, is that there are continuously new ideas for how to conduct research, and there has to be priorities made concerning to which paths to choose. It is still important to consider every idea and approach, because in a way it gives us a better understanding of the big picture. The ideas may be worth documenting, both the idea in general and the essence of the idea, as this is something that can be used in future work which may include even more information.

5.4 Reflection on the process

It is important to note that these different approaches do not answer the same question of similarity, because they each have a unique way of representing results. There is also an important question on what should have been the approach from the beginning, and the positive effects of this would be that we would have gained the results on an earlier stage, and we might also have learned a lot more from this approach. On the other hand, going straight into a broad approach, would mean that we do not really know anything of the things we have learned along the way going into this. By choosing a combination of approaches, the results have more depth and meaning for us. The path with moving from a narrow approach to a broad one, have contributed to a better understanding. This was not the intended approach to begin with, but it has still shown to be a better approach for understanding, and has not had any bad consequences in regards to the project.

What would the alternatives have looked like if only one of the two presented approaches was chosen? If we had gone directly into the narrow approach, the results would have been achieved a lot faster, and given us room to further developed the tools in this project. On the other hand, if we went straight into a broad approach, we would have gained the matrix with all the information about all video games involved in the data set, a lot faster. This again, would have given more room for testing other methods of correlation, for instance Point-biserial, which as Pearson's R is a linear correlation coefficient. Other methods that could have been tested is Cramér's V or Kendall's tau, which as non-linear correlation methods, just like Spearman's Rho.

If we had chosen only one of these approaches, we would have ended up on a unfortunate situation, because the narrow approach would not present us with the opportunity to conclude or form a credible opinion which would be generalizable. On the other side, only going for the broad approach would have put us in a situation where we have all the answers, but we would not have known anything about them or what they actually mean.

The 2000 applications in the dataset are the applications in Steam with the highest player count during the last 24 hours on the 11th of December 2017, 1963 of these are actual video games. During this research, all applications involved in the dataset has been a part of the calculations.

Amongst the 25 highest correlating video games, we find that there a lot of strategy based video games, which are big and can be time consuming to play, which again leads to the ability to play over a longer period of time. In addition to strategy games, there are some exceptions which also are quite popular in this period of time.

Counter-Strike and Football Manager, are, for example video games that one might play for several hours at the time. Although these games are quite different, due to the reason that Counter-Strike is a more intense game that requires a lot of constant attention, while Football Manager is a video game with a bit more slow pace, that can be running whilst doing other things.

There are two different phenomenon's found in these results, where one of them is related to succession within a video games series, or version likeness in general. Succession in this case means that a particular video game, or video games are a part of a bigger franchise or a part of a series where there are more than one game with the same setting. Both predecessors and successors can be found as a combination in the 25 highest correlating video games, both from part 1 and part 2 of the player count history. We also find variants or DLC's within these two tables, a DLC is additional content for an already existing and published video game, that is provided by the published of the game itself, and it provided through the internet. This is not intuitive and it does not tell us anything at all in regards to the concept of similarity, because DLC's are a part of an already existing game. This particular occurrence is presented in Table 4.4 on line 17.

It is to a degree strange that there are so many video games from the same franchise that have a high correlation value, nearly a perfect correlation, in one way it makes sense because well-established franchises often have a loyal fanbase. Only one of the combination of video games found in the highest 25 correlating video games, from both parts, correlates with a game that is completely different from itself. This is

found on line 18 in Table 4.6 Apart from that, it might be necessary to look further into the results to find something that is more “different”.

The fact that there are multiple games from the same franchise is not necessarily surprising, because they are, after all, in the same series, but these results does not represent what we were interested in when it comes to similarity. Moving forward with PCA on these results, would also result in a misguided impression of what similarity in video games is.

The lowest correlating video games from part 2 of the dataset, as found in Table 4.7, show that *Kenshi, Gloria Victis* and *Ragnarok Clicker* appear multiple times in the results. Starting with Kenshi, it is possible that this video game have a very local following, possibly in Asia, which again leads to the amount of players being higher during a different time of day than a lot of other games, which again leads to the pattern of the game being “different” from others, and the result of that is a low and/or negative correlation. Moving forward, we also have Gloria Victis which have few players, and it is possible that this video game also is extremely popular in one particular place. Last of the three forementioned games, we have Ragnarok Clicker, and a trait of clicker games is that it is a game that can be running over a long period of time without needing a particularly intense focus, so it is possible that this game is being “played” while begin at work or school.

When it comes to Football Manager, it is possible that they have such a high correlation because of the continuous new versions of them game. There is a possibility for several things to explain the high correlation within these games, one of these factors can be that these games are have an extremely loyal fan base, and some people might play the older versions simply because they enjoy the games. In addition, they might be played less and less, but there is a possibility that they have the same curve when it comes to player count, and therefore still appear as similar in this research because they numbers in general are quite weak. Meaning that all of the video games in the Football Manager are possibly quite popular, especially in the following time after a release of a new game. Therefore, all these games will have a higher number of players in the beginning, while the pattern when it comes to player in these games will slow down and flat out as a new game is being released.

The main takeaway from the top results is that video games that basically are the same game, is dominating the top of the list. This does not tell us anything new about video games, because this is something that we could have guessed beforehand, both because of knowledge of video games but also just by looking at the titles in a particular franchise. There is a big group of video games from franchises that dominates the result, which leads to a problem in the pursuit of similarity. The dataset is of such a long period, and this leads to franchise games having the time to be replaced, and therefore a shorter timeframe would not show this phenomenon. These games dominates our results, and will probably also dominate a further examination with the help of clustering methods, as the mentioned PCA. For future research it is important to note these findings, and in additions it provides better foundation for the future, because we have more awareness surrounding the dataset and knowledge related to it.

It is also difficult to determine what high correlation actually is, but a magical limit is often above 0.7 or 0.8, for it to count as high correlation. In this case, a correlation coefficient above 0.6 can also be sufficient, but there is no general idea of what is actually good enough. It is also not straightforward to know what is needed in order to get a desired result.

The results in this project provides more information about the applicability of datasets, and the factors that are beneficial to be aware of when it comes to a research project. We could have dug deeper in order to gain more information beforehand, but it would still be difficult to know which difficulties we would run into before going into the project itself. This is mainly in regards to the domination of franchise games, and the questions surrounding what could have been done differently for these particular values. For instance, the values could have been summed up, but that would potentially lead to a higher number than the newest version in the franchises, which would be misleading.

5.5 Reflection on assumptions and hypotheses

The assumptions and hypotheses created for this research have been an interesting part of this project, as they contribute to gaining more knowledge about the patterns and how they behave. All of them have not been verified, and some have been partially verified, but this is something that can be done in future research. In retrospect, it may have been beneficial to include an assumption that states something in relation to how well a similarity function actually works.

5.6 Discussion on similarity

What defines if a game is similar to another or not? There will almost always be some presumptions when it comes to the subject of determining similarity. The dataset in our research is based on the amount of players at a certain time, and therefore the question is also if this is actually something that can determine similarity or not. When it comes to video games there are several different things we can look at and these aspects allow us to make up our minds and opinions for ourselves. This can be the genre of the game, what kind of game it is when it comes to playing style, or the perspective of the game. It is an obvious assumption to say that Total War: Warhammer II and Age of Empires are similar because they are real time strategy games. Real time strategy are games where the players do not play in turns, but play simultaneously. The contrast to this are turn based strategy video games. Another example is if we were to say that Dead Space and Resident Evil are similar games, because they are games of the same genre, which is horror. Even the fact that a game is published or developed by the same company, can be something that makes video games similar in someone's opinion.

The question still is if similarity can be based on the player count itself, and therefore it is interesting to conduct some research into games we assume are similar, even if that is something that turns the research into a certain direction due to presumptions. Moving forward it is important to factor in outside factors that can affect the results in some way. These factors can both be things that are not related to video games in any way, and some things that might be somewhat related.

For instance, it is important to check different dates, and see if the numbers and correlation changes based on these factors. The meaning of this is that there are several different holidays during the year, like Christmas, Easter, constitution days, Black Friday and other national or international holidays. It is natural to spend more time with family and friends around these times, or maybe to do something entirely different on these particular days. If we also assume that the average person that plays video games works a normal eight to four or nine to five job or is at school, the numbers will probably look different in the evenings during the week and during the weekend, especially Friday night and Saturday. This is something that is purely based on what we relate to what we know as a “normal” work schedule. Furthermore, the weekends are a time where a lot have time off from work, and for many people gaming is a social activity, which is something that might be easier to plan for during the weekend. By social activity, the meaning is that a lot of video games are games that can be played with others, both in person or over the internet via a communications platform, for instance Discord or TeamSpeak.

In the video game industry there are also different events and happenings, often something that is arranged annually or even multiple times during a year. A lot of different games, like Dota 2, Fortnite, Counter-Strike: Global Offensive, League of Legends and many more, have tournaments, where there are many contestants. This is something that is interesting to look into, in order to figure out if something changes with the data from the game in that particular time. Changes can be related to if there are more players as a result of engagement for the game in particular, or there might be less players because people simply want to be a bystander and follow along the tournament.

There are also different game conventions, for instance BlizzCon which is held by Blizzard Entertainment, and TwitchCon which is for the livestreaming videogaming platform Twitch. These are also something that might factor into the numbers on the particular dates when these events occur, and therefore it can be something that affects the outcome. Furthermore, there are also computer festivals or LAN’s (local area networks) being arranged, where the biggest one is DreamHack. DreamHack in several different countries on different date, and their headquarters are located in Sweden. As with all the different events that are mentioned, this might affect game

time on a particular date or dates, and it can be both higher or lower number of players.

Another aspect that may have an impact on numbers and when it comes to similarity is all the recent developments and new features within the video game industry, like subscriptions to different services. For instance, Xbox Game Pass and PC Game Pass by Microsoft, which are subscription based and offers a selection of video games that you can play if you have an active subscription. These services offers a rotation of video games, which provides the players to try different games continuously. This is not related to Steam which the dataset in this project is from, but it is still an interesting aspect of video games and the industry, which is an important factor that may affect similarity, and therefore, is an interesting point for future research within this subject.

We are currently in a digital age, where there is continuously new development, releases and news happening, especially in the video game industry. As with many other mediums, for instance, tv shows, movies and books that have been mentioned, there is countless options and might also be difficult to know what to look for or what to prioritize when deciding to try something new. As with new services that revolves around subscription, there is also a question of whether we own or lease digital entertainment. This is not necessarily directly related to similarity, but it can have an impact on what games an individual chooses to try out, and therefore might also impact the similarity in this research. The impact can be because one choses to try out a lot of different video games, instead of staying loyal to one particular video game, and it can also affect the number at a certain point because of new releases. New releases in general can influence player count, if only for a short amount of time, as individuals might want to try a new game that is released in a genre that they prefer, but end up going back to a more popular or stable game that they know they enjoy.

As mentioned national and international holidays can affect the numbers, and we have recently experienced a global pandemic, which is something that in many ways also can affect the numbers in the dataset. As the dataset includes player count from the 14th of December 2017 to the 12th of August 2020, therefore it can be interesting

to compare the numbers from before 2020 with the numbers from early 2020 leading up to the last date represented in the dataset. The reason for this is that a lot of people were temporarily laid off, in addition to a lot of companies that had to declare bankruptcy as it was a difficult period of time and hard to make it due to all the different restrictions in the society that happened globally. That again, leads to reason to believe that some people spent more time playing video games, or there might even have been new people that were introduced to video gaming, and something that a lot of people used to be able to do something socially, even if it was online. This is just an assumption, and would not necessarily contribute to any more knowledge of similarity within video games, but can put in context with the fact that a lot of people took up new hobbies and started doing new or different things during the pandemic.

5.7 Similarity in video games in context with cloud computing

One problem area related to this is cloud gaming, or gaming-as-a-service, which is online gaming where video games run on remote servers, where the video game is streamed directly to a user's device or played remotely from a cloud. This requires significant infrastructure to work as excepted, and needs data centers and server farms for running these games. There is therefore a question of how video games and services revolving around them, are connected to this project. Additionally, there is a service problem related to the correct allocation and management of resources. Some of the findings in this research can to a certain extent be resourceful when it comes to the allocation of resources, as the subject of similarity gives us more information about certain video games. Something that is especially interesting is the video games that have a large player base, as we know that these will demand more capacity and virtual machines for running in the cloud. It is important to note that a high or positive correlation does not lay a foundation for resource allocation or management in itself, but it can provide better insight as to how much and when a video game is played the most.

When it comes to how to predict and allocate the necessary resources for cloud operations, it can be beneficial to both compare a video game to itself and others,

because it would give more information. Both because we can learn a lot from the video game data from one game itself, but the perspective can be broadened when looked at in comparison to others. These are factors that can be used for this particular aspect of information technology in general, because the result of having more data is having more knowledge about a certain video game. Even though it might be beneficial to have even more in-dept information, and to dig deeper within some of the findings, for instance by looking at certain dates, longer time periods or a certain time period during the day. This, in combination with other known facts surrounding a video game, for instance when a new DLC is being released, helps us prepare and manage better. This is also something that can be related to a new game being released, especially if we on beforehand know that a new released is related to something that we have already found to be similar.

The results in this project can be benefited from by video game companies that have multiple video games. One of the reasons for this is that it gives more insight, for instance if one their video games are more popular than others. It can also be beneficial for video game companies with franchises, as these game often have the same behaviour, especially if it is a well-established franchise, as we can see with the Football Manager phenomenon in Table 4.4. When it comes to these types of video games, their lifecycle is important because they as mentioned behave the same way and with a loyal fanbase there will always be someone playing the older versions, even when a new version is released. The newest version will probably always be the one that have the most players simultaneously and over a period of time, but there might be other reasons that a person does not play the newest version, for instance in general being a fan of a particular edition of the franchise or maybe not buying the newest straight away.

When it comes to the case of a video game developer having one or more games, we still cannot count on them being similar even though we might think that they are. In certain video games, we might find that they are very similar with a strong correlation a certain point, maybe even for weeks or months in a row, but this might not always be the case. If these numbers are not looked at in a long perspective, the case might be slightly different as well. As stated earlier the time period of the dataset used in this project, is approximately 2,5 years, and when the timespan is so

long, it makes it more difficult to form an opinion when it comes to day-to-day operations. This leads to it being a time-consuming job, where we do not know what happens from one day to another, which makes it unpredictable. Video games are different from other types of services, in general, because it is mainly a leisure activity.

These video games will not necessarily have the same pattern, also because the video games people choose to play continuously changes. The changes in video games to play is somewhat dependent on the player itself, because some are loyal to one particular game or franchise, whilst others might enjoy trying different video games both in a variety or genres, or within the same genres. Recent developments within the video game industry also includes services like Xbox Game Pass and PlayStation Plus, where you can have a monthly subscription and get access to an assortment of different games, that are included in these services. This is not something that is relevant in the context of the particular dataset in this project, as this data were retrieved a few years ago. This is still an important fact to note for future research, because these are services that are more and more used as it provides a better selection for video game players and might broaden their horizon.

A possible future direction for this research is a comparative study, which is what Jon-Erik Tyvand (2011) did in his master thesis, through a predicative algorithm tested on Counter-Strike: Source, Football Manager 2010 and Supreme Commander 2. This revolves around a video games data in itself, not being compared to another [67]. The method in this type of approach would be to train first, and then act on the takeaway of what we have learned. A question within this type of method could be how a video games own data compares to other games that are assumed to be similar. The meaning behind this is to look at something other than the video games selected for this being in the same genre, and it cannot be drawn randomly, and it comes down to a video games own history versus history from a different video game. Presumably a video game is more similar to itself yesterday, than it would be similar to a different video games data from yesterday.

Chapter 6 Conclusion

Similarity is an ambiguous concept, as it can mean different things to different people. For instance, with the sunset similarity related to franchises, it could be correct to assume that most people would classify the Football Manager video games as similar just based on the name. There is an underlying question of what similarity actually is, as it can be different from one person to another, and one's opinion does not make up the truth about what similarity is. In general, the video games that were selected for the narrow approach in this project, appeared as similar in the cases were we expected similarity. Although they did not have as high of a correlation value as the franchise video games that were found to have a very high correlation when the broad approach were conducted. These games were also influenced by how the data were split into intervals, and some of them had a very high correlation on certain dates. This is a promising method of continuing research on this dataset, but it needs more work to refine the intervals. Similarity has been found, but it is not bulletproof.

There is a need to conduct an examination of more than one method, in other to understand the concept of similarity within video games. This is because we do not know enough about them to understand how they behave. As Begnum and Burgess (2004) concludes in their paper above anomaly detection, "there does not seem to be a compelling argument for central analysis of the patterns of behaviour", even though the themes of this project is quite different, it can be seen in context with this. The reason for that is that we would not get anywhere with only the broad approach, as it would be a lot of correlation coefficients, which little knowledge about them. This project has become more and more clear throughout the process, and therefore the tools and the path of the research have become clearer as a result of that. For instance, the fact that some of the files belonging to a certain video games, might have missing values or 0 as a value. This is something that could have been changed in this dataset, by changing those values in order to avoid having them in the dataset in general. This is something that would change the meaning of the data,

because if we lack the particular values, it would not necessarily tell us anything different by changing them ourselves.

Some of the phenomenon’s presented in this project makes it apparent that similarity cannot be looked at as an isolated subject, without having more categorization surrounding certain aspects related to the video games. There are certain phenomenon’s that are distinctive for video games themselves, that confirm our assumptions, but genre and time is not a guarantee for similarity. This opens for further questions, which will be presented in the next section.

6.1 Future work

This research opens up for several different ways to continue the pursuit of similarity within video games, and as mentioned it is a big industry with a lot of different factors that can affect results in one way or another. One of these questions is, why are video games that should be similar, not similar after all? There can be a lot of different factors that affect this, for instance that local fanbase influence more than genre itself.

Reference List

[1] World Economic Forum (n.d.) *Fourth Industrial Revolution*. Retrieved 14.05.20, from: https://www.weforum.org/focus/fourth-industrial-revolution

[2] Sweney, M. (2021) *Facebook outage highlights global over-reliance on its services.* Retrieved 12.10.21, from: https://www.theguardian.com/technology/2021/oct/05/facebook-outage-highlights-global-over-reliance-on-its-services

[3] Beattie, A. (2021) *How the Video Game Industry Is Changing.* Retrieved 12.10.21, from: https://www.investopedia.com/articles/investing/053115/how-video-game-industry-changing.asp

[4] Brush, K. (2022) *data visualization.* Retrieved 14.05.23, from: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

[5] Rubi, O. (2018) *What can Big Data do for BI?* Retrieved 11.05.20, from: https://www.clearpeaks.com/what-can-big-data-do-for-bi/

[6] Trepte, S., & Reinecke, L. (2010). Avatar creation and video game enjoyment. *Journal of Media Psychology*.

[7] Sartorius (2020) *What Is Principal Component Analysis (PCA) and How It Is Used?* Retrieved 15.10.21, from: https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186

[8] twinkl (n.d) Video Games. Retrieved 15.10.21, from: https://www.twinkl.no/teaching-wiki/video-games

[9] We Pc (2021). *Video Game Industry Statistics, Trends and Data In 2023. Retrieved* 14.05.23, from: https://www.wepc.com/news/video-game-statistics/

[10] PC Games for Steam (n.d.) *What is Steam?* Retrieved 10.10.21, from: https://pcgamesforsteam.com/what-is-steam

[11] Merriam-Webster (n.d.) *guild.* Retrieved 16.10.21, from: https://www.merriam-webster.com/dictionary/guild

[12] Bråten, O. (2019) *På tide å skru av «Fortnite»?* Retrieved 14.05.23, from: https://www.aftenposten.no/meninger/debatt/i/WbJ8XK/paa-tide-aa-skru-av-fortnite-ole-andre-braaten

[13] Johannes, N., Vuorre, M., & Przybylski, A. K. (2021). *Video game play is positively correlated with well-being. Royal Society open science*, *8*(2), 202049.

[14] Barnekreftforeningen (2020) *Norske gamere satte innsamlingsrekord for Barnekreftforeningen.* Retrieved 31.10.21, from: https://www.barnekreftforeningen.no/nyheter/gamere-mot-barnekreft

[15] Twitch (n.d.) *Nerdelandslaget.* Retrieved 04.04.23, from: https://www.twitch.tv/nerdelandslaget

[16] Bajpai, P. (2021) *How Cloud Computing is Changing the World of Games and Gaming.* Retrieved 30.10.21, from: https://www.nasdaq.com/articles/how-cloud-computing-is-changing-the-world-of-games-and-gaming-2021-02-24

[17] Amazon (n.d) luna. Retrieved 14.05.23, from: https://www.amazon.com/luna/landing-page

[18] Microsoft (n.d.) Project xCloud. Retrieved 14.05.23, from: https://developer.microsoft.com/en-us/games/products/project-xcloud/

[19] Moss, K. (2018) Vi kunngjør Project Atlas. Retrieved 14.05.23, from: https://www.ea.com/nb-no/news/announcing-project-atlas

[20] Di Domenico, A., Perna, G., Trevisan, M., Vassio, L., & Giordano, D. (2021). A network analysis on cloud gaming: Stadia, GeForce Now and PSNow. *Network*, *1*(3), 247-260.

[21] Compañ-Rosique, P., Molina-Carmona, R., Gallego-Durán, F., Satorre-Cuerda, R., Villagrá-Arnedo, C., & Llorens-Largo, F. (2019). A guide for making video games accessible to users with cerebral palsy. *Universal Access in the Information Society*, *18*, 565-581.

[22] Jaliaawala, M. S., & Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey. *Artificial Intelligence Review*, *53*(2), 1039-1069.

[23] Rouse, M. (2019) *Big data*. Retrieved 04.06.20, from: https://searchdatamanagement.techtarget.com/definition/big-data

[24] Barr, P., Noble, J., & Biddle, R. (2007). Video game values: Human–computer interaction and games. *Interacting with Computers*, *19*(2), 180-195.

[25] Fisher, D. et al. (2012) *Interactions with Big Data Analytics.* Retrieved 04.06.20, from: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/inteactions_big_data.pdf

[26] Optimizely (n.d.) *A/B Testing*. Retrieved 06.06.20, from: https://www.optimizely.com/optimization-glossary/ab-testing/

[27] Elmqvist, N., & Irani, P. (2013). Ubiquitous analytics: Interacting with big data anywhere, anytime. *Computer*, (4), 86-89.

[28] Haddadi, H. (2016). Human-data interaction. *Encyclopedia of Human Computer Interaction*.

[29] Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2014). Human-data interaction: The human face of the data-driven society. *Available at SSRN 2508051*.

[30] Rouse, M. (2019) *user interface (UI)* Retrieved 06.06.20, from: https://searchapparchitecture.techtarget.com/definition/user-interface-UI

[31] DeveloperSpace (n.d.) *What are Adaptive User Interfaces?* Retrieved 06.06.20, from: https://ds.gpii.net/content/what-are-adaptive-user-interfaces

[32] Yigitbas, E., Jovanovikj, I., Biermeier, K., Sauer, S., & Engels, G. (2020). Integrated model-driven development of self-adaptive user interfaces. *Software and Systems Modeling*, 1-25.

[33] Alvarez-Cortes, V., Zárate, V. H., Uresti, J. A. R., & Zayas, B. E. (2009). Current challenges and applications for adaptive user interfaces. *Human-Computer Interaction*, 49.

[34] Gullà, F., Ceccacci, S., Germani, M., & Cavalieri, L. (2015). Design adaptable and adaptive user interfaces: A method to manage the information. In *Ambient Assisted Living* (pp. 47-58). Springer, Cham.

[35] Wannigamage, D. et al. (2020) Steam Games Dataset: Player count history, Price history and data about games. Retrieved 10.10.21, from: https://data.mendeley.com/datasets/ycy3sy3vj2/1

[36] Statistics Solutions (n.d.) *Correlation (Pearson, Kendall, Spearman).* Retrieved 09.11.21, from: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/correlation-pearson-kendall-spearman/

[37] Statistics Solutions (n.d.) *Conduct and Interpret a Spearman Rank Correlation.* Retrieved 09.11.21, from: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/spearman-rank-correlation/

[38] CR, A. (2023) *Exploring Clustering Algorithms: Explanation and Use Cases.* Retrieved 14.05.23, from: https://neptune.ai/blog/clustering-algorithms

[39] Jaadi, Z. (2021) *A Step-by-Step Explanation of Principal Component Analysis (PCA).* Retrieved 10.10.21, from: https://builtin.com/data-science/step-step-explanation-principal-component-analysis

[40] Begnum, K., & Burgess, M. (2005). Principle components and importance ranking of distributed anomalies. *Machine Learning*, *58*, 217-230.

[41] SteamDB (n.d.) *Most played games on Steam.* Retrieved 14.05.23, from: https://steamdb.info/graph/

[42] Begnum, K., & Burgess, M. (2007). Improving anomaly detection event analysis using the eventrank algorithm. *Lecture Notes in Computer Science*, *4543*, 145.

[43] Floros, G., & Siomos, K. (2012). Patterns of choices on video game genres and Internet addiction. *Cyberpsychology, Behavior, and Social Networking*, *15*(8), 417-424.

[44] Björk, S., & Holopainen, J. (2005). Games and design patterns. *The game design reader: A rules of play anthology*, 410-437.

[45] Hodges, B. (2023) The Last of Us is the latest video game adaptation plagued by the same problem. Retrieved 14.05.23, from: https://www.polygon.com/23645114/last-us-tv-show-game-comparison

[46] Hilgard, J., Engelhardt, C. R., & Bartholow, B. D. (2013). Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (GAMES). *Frontiers in psychology*, *4*, 608.

[47] Web Archive (2018) *The Gathering 2018.* Retrieved 09.11.21, from: https://web.archive.org/web/20180807185944/https://www.geekevents.org/tg18/

[48] Teachoo (2021) *What is reflexive, symmetric, transitive relation?* Retrieved 09.11.21, from: https://www.teachoo.com/7061/1160/What-is-reflexive--symmetric--transitive-relation-/category/To-prove-relation-reflexive--transitive--symmetric-and-equivalent/

[49] MasterClass (2021) *Battle Royale: A Guide to Battle Royale Video Games.* Retrieved 10.11.21, from: https://www.masterclass.com/articles/what-is-a-battle-royale

[50] pcmag (n.d.) *streaming video games.* Retrieved 11.11.21, from: https://www.pcmag.com/encyclopedia/term/streaming-video-games

[51] socialblade (n.d.) *Top 100 most followed twitch accounts (sorted by followers count).* Retrieved 11.11.21, from: https://socialblade.com/twitch/top/100

[52] Total War: Warhammer Wiki. (n.d.) *Total War: Warhammer II.* Retrieved 08.02.23, from: https://totalwarwarhammer.fandom.com/wiki/Total_War:_Warhammer_III

[53] Wikipedia (n.d.) *Cities: Skylines.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Cities:_Skylines

[54] Wikipedia (n.d.) *Civilization IV.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Civilization_IV

[55] Dead Island Wiki (n.d.) *Dead Island: Riptide.* Retrieved 08.02.23, from: https://deadisland.fandom.com/wiki/Dead_Island:_Riptide

[56] Official Dying Light Wiki (n.d.) *Dying Light.* Retrieved 08.02.23, from: https://dyinglight.fandom.com/wiki/Dying_Light

[57] Wikipedia (n.d.) *Payday 2.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Payday_2

[58] Wikipedia (n.d.) *Tom Clancy's Rainbow Six Siege.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Tom_Clancy%27s_Rainbow_Six_Siege

[59] Wikipedia (n.d.) *Counter-Strike: Global Offensive.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Counter-Strike:_Global_Offensive

[60] Wikipedia (n.d.) *Counter-Strike.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Counter-Strike_(video_game)

[61] Wikipedia (n.d) Video game modding. Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Video_game_modding

[62] Wikipedia (n.d.) *Football Manager 2017.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Football_Manager_2017

[63] Wikipedia (n.d.) *NBA 2K18.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/NBA_2K18

[64] Wikipedia (n.d.) *Final Fantasy XIV.* Retrieved 08.02.23, from: https://en.wikipedia.org/wiki/Final_Fantasy_XIV

[65] Wikipedia (n.d.) MechWarrior Online. Retrieved 03.05.23, from: https://en.wikipedia.org/wiki/MechWarrior_Online

[66] Lewis, R. (n.d.) Is it possible to learn a language playing video games? Retrieved 14.05.23, from: https://blog.pimsleur.com/2020/07/08/learn-a-language-playing-video-games/

[67] Tyvand, J. E. (2011). *On the predictability of server resources in online games, an investigative approach* (Master's thesis).