NOVA
IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

ASSESSING COVID-19 IMPACT ON USER OPINION TOWARDS VIDEOGAMES

Sentiment Analysis and Structural Break Detection on Steam Data

Pedro Nuno Ângelo Mota

Dissertation presented as requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

ASSESSING COVID-19 IMPACT ON USER OPINION TOWARDS VIDEOGAMES

SENTIMENT ANALYSIS AND STRUCTURAL BREAK DETECTION ON STEAM DATA

by

Pedro Nuno Ângelo Mota

Dissertation presented as requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence and Knowledge Management.

Co-advisor: Professor Doutor Flávio Luís Portas Pinheiro

Co-advisor: Professor Doutor Bruno Miguel Pinto Damásio

NOV 2021

ABSTRACT

As we live in a world where the videogame industry grows day by day and new media is constantly emerging, user feedback can be widely found online. User reviews are a highly valuable data source when studying player perception of a videogame. They are also apparently volatile to updates released by developers and other external events, which may change user opinion over time. Here we assess whether the COVID-19 pandemic outbreak fell in this category, having or not a noticeable impact on the player view and popularity of videogames.

In this research, we build and implement a method to collect active player data and user reviews, identifying the sentiment contained in the expressed opinions. Furthermore, we investigate the existence of structural breaks in the time series we target. For this purpose, we targeted user-reviews and active player data collected of Steam's twenty most popular Massive Multiplayer Online Role-Playing Games. To collect sentiment polarity values, two Natural Language Processing Python libraries were used, TextBlob and VADER, and structural break detection was put into practice using strucchange R package.

The results of this work show us that despite having a great effect on the number of active players, the COVID-19 pandemic did not produce the same impact on Steam user reviews. Nonetheless, we were able to identify one of the platform's major reviewing related updates as a structural break. We believe this approach can be used for further assessments on public opinion towards a specific product, in the future.

KEYWORDS

Videogames; User Reviews; Sentiment-analysis; Structural breaks; COVID-19;

INDEX

FIGURE INDEX

TABLE INDEX

LIST OF ABBREVIATIONS AND ACRONYMS

API Application Programming Interface – Set of functions and procedures allowing the creation of applications that access the features or data of an operating system, application, or other service.

DLC Downloadable Content – Additional content created for an already released videogame, distributed through the Internet by the game's publisher.

ML Machine Learning – A part of Artificial Intelligence that automates analytical model building.

MMORPG Massively Multiplayer Online Role-Playing games – Online games with large numbers of players, often hundreds or thousands, on the same server.

NLP Natural Language Processing – Subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages.

NLTK Natural Language ToolKit – Python library that contains programs and libraries used for natural language processing.

POS Part of Speech – The eight parts of speech in the English language: noun, pronoun, verb, adjective, adverb, preposition, conjunction, and interjection.

SVM Support Vector Machines – Supervised learning models used for regression, classification and outliers detection.

TF-IDF Term-Frequency and Inverse-Document-Frequency – A statistical measure that assesses how relevant a word is to a document in a collection of corpus.

VADER Valence Aware Dictionary and Sentiment Reasoner – A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

1. INTRODUCTION

Written reviews can be found online, regarding all kinds of products and software. Videogames are no exception. The vast amount of reviewing and social media platforms, hold a massive amount of user experiences and evaluations. While the interest to analyse this type of data has been increasing day by day, still a lot can be done in this area. Public written reviews, unlike professional ones, are often more polarized, easier to understand and frequently written a long time after a videogame release (T. Santos et al., 2019). As a result, these reviews are susceptible to be affected by major updates and other large-scale events that may occur over time. This creates an opportunity to study how those events may impact the user experience while playing videogames, by considering the date in which they were written. Such analysis would provide valuable insights for videogame developers and the industry in general, who seek to improve user experience.

Figure 1 – COVID-19 outbreak indirect impact on videogames

As of present time, the COVID-19 outbreak is the most relevant example of an impactful worldwide event, leading to a dramatic shift in our day-to-day habits (Chakraborty & Maity, 2020). In particular, many countries implemented social distancing and confinement measures that introduced dramatic consequences to our social lives (Farboodi et al., 2020). For instance, we moved our social interactions to the internet, exchanging casual social interactions by virtual ones (Farooq et al., 2020). In post-pandemic world, technological companies have capitalized in these new behaviours, with an increasing demand for new entertainment digital content. Computer and console games have also seen a surge in activity in the aftermath of COVID-19. In fact, Steam, a video-game distribution platform registered a 23.8% increase in the amount of concurrent users between the first and the last day of March 2020 (Steam Database, 2020). A resume of this situation can be seen in Figure 1.

As a result of COVID-19, many individuals reported poor mental health and higher negative mood during lockdown. Negative psychological effects include post-traumatic stress symptoms, confusion, and anger (Brooks et al., 2020). Causing factors variate from experiencing stressful situations due to the risk of catching the virus, to deprivation of seeing relatives or close friends or even experiencing bereavement. Additional stressors include long quarantine duration, frustration, boredom, inadequate supplies, inadequate information, financial loss, and stigma. Furthermore, certain lifestyle aspects may have been changed, such as diet, sleep quality and physical activity (Ingram et al., 2020). Regarding the role played by videogames, recent studies point out that many people have found a way to connect with friends and other people through playing video-games, improving their own social integration and connectedness (Marston & Kowert, 2020; Riva et al., 2020). Here we ask whether such effects on individual mental health reflected in the sentiment they reveal towards their experience in videogames.

It is widely known how social media serves as a unified platform for users to express their thoughts on all kinds of subjects, including videogames (Hu et al., 2017). There are however more ways for players to express their feelings and considerations towards videogames. One of the main options is to write a review on websites created for that same purpose i. e. Metacritic, or in videogames distribution platforms which include that feature i. e. Steam. Game reviews among other aspects, often include first-person opinions of the experience the game reviewer had with the game in question. In many cases, they consist of descriptions of the emotions felt both during play as well as after (Zagal et al., 2009).

Past studies about sentiment analysis on videogame user reviews include the analysis of in-game player chat messaging (Thompson et al., 2017), assorted game related tweets (Islam et al., 2016) and written reviews. A fair number of studies have been conducted on videogame amateur reviews in order to extract sentiment. A previous work from 2019 includes a review on the existent studies by then (Rajapakshe, 2019). This review detects two main types of approaches, binary classification and aspect-based classification. Binary classification methods evaluate if the sentiment displayed is positive or negative and compares with whether the user recommended or not the game. Aspect-based classification provides more intuition about the feedback when compared to binary classification, allowing for example, to separately study feedback on certain features. Recent studies also elaborate how COVID-19 impacted user experience on Location Based Games, demonstrating the influence that this type of games had on human behaviour during this crisis (Laato, Islam, et al., 2020; Laato, Laine, et al., 2020).

However, there is still a lack of research on the impact of large-scale events and major updates on user experience feedback and close to none approaches were found of structural break test applications on videogame related data. Regarding the COVID-19 pandemic, no approaches have been found on detecting its general impact on videogames by analysing user reviews. This study is expected not only to deepen the knowledge regarding user sentiment analysis, but also to detect structural breaks in a new context and investigate further impacts of the COVID-19 pandemic in the videogames sector. To do so, we focus on detecting the impact of large-scale events such as the COVID-19 pandemic and major game or platform updates on videogames. This includes assessing the sentiment displayed in user reviews, as well as the total amount of reviews and the numbers of active players over time. In order to achieve the defined goal, user reviews and active players data from recent years will be collected and processed. Sentiment polarity values will be extracted from the reviews using a set of algorithms. Lastly, we will pick a set of variables from the data to be studied and define a model to test for the existence of structural breaks. This procedure is expected to provide us the means to understand whether the major changes can be associated with the impacts caused by the pandemic or other major events we expect to identify.

The structure of this work is organized as follows: Theoretical revision, where a summary of existing literature will be presented and divided in videogames, Sentiment Analysis and Structural Break Tests and also where the basis of our method will be defined. Results and Discussion section includes our procedure in order to collect and clean the data, extract sentiment polarity and conduct structural break tests. The obtained outcomes, along with a critical review of the findings complete this section. The last part is the conclusions obtained from this work, as well as the identification of limitations and future work.

2. THEORETICAL REVISION & METHOD

In this section we begin to discuss how the videogames industry has grown over recent years and the currently available videogame opinion data sources. Afterwards, we explore the available sentiment analysis and text mining techniques by assessing existent studies in these areas. Lastly, a review on methods to detect structural breaks in time series and how they can be applied in this scenario.

2.1. VIDEOGAMES

We live in a digital era, where the usage of the internet by consumers for product purchases increases day by day. The videogame industry is growing considerably, totalling $177.8 billion in 2020, a 19% increase from the yearlong revenue in 2019, also boosted by the COVID-19 lockdown measures (Newzoo, 2020). The huge revenue and user base all over the world, characterize this industry as one of the largest and most competitive entertainment markets and user experience plays a key part in it. User empowerment allows producing companies to obtain a clear view of customer needs and reactions to product release. Through positive or negative feedback new opportunities can be identified to improve user acceptance and polishing new launching ideas (González-Piñero, 2017).

User feedback can be found online in large quantity, in many different places. Social media are a valuable source of opinions on all kinds of products, including videogames. When using Twitter, a simple search for the desired product could display an enormous quantity of related expressed thoughts and feelings in the form of very short texts. Previous studies include the extraction of game related tweets using the free version of the official Twitter API, which has the disadvantage of restricting data collection to the past 7 days (Bonenfant et al., 2020). In order to avoid this issue, a script can be created using a Python Library to scrape data directly from Twitter. Such method enables the extraction of a given number of tweet within a user-defined time period (Roy et al., 2018). The downside of studying Twitter data, in the context of our study is that tweets tend to be mostly neutral, as not many express either positive or negative sentiment unlike actual reviews that express opinions (Hagen et al., 2015).

Metacritic is an example of a popular website created specifically to aggregate reviews and ratings of diverse entertainment goods such as movies and videogames, both from expert critics and regular users (T. Santos et al., 2019). Reviews contain a simple text input field and a score from 0 to 10, given by the author and display the number of times it was marked as helpful by other users. Every user registered in the platform can pick any game to review, despite owning the game or not (Kasper et al., 2019). Past works include the extraction of reviews and scores from Metacritic using an open source crawler for Java, as well as review indexing and data cleaning (Ruseti et al., 2020). Another work used two different scripts to extract reviews, a scrapper script and a python API to collect data from Metacritic, amongst other sources. (Quader et al., 2017).

Figure 2 – Steam review example

Steam is a multi-billion dollar distributed gaming platform that acts as a third-party medium to sell games online and download them (Bais et al., 2017). One of its features is allowing users to write their own reviews, assigning positive or negative ratings to any game. Large numbers of user reviews are written every day and displayed on the respective game's store page, providing a valuable help for other users who are looking to decide whether or not to buy that game (Lin et al., 2019). The review in Figure 2. provides an example of a Steam user review and how it is displayed on their website. Although Steam Reviews do not contain a concrete rating but a positive or negative recommendation instead, they contain other useful features and only users who bought the game can leave a review. In this manner, we will collect data from Steam for the purpose of our study. Considering we want to study reviews until as recent time as possible, we will be collecting data ourselves, instead of using an already existent dataset.

Regarding the target videogames for our study, we will focus on Massively Multiplayer Online Role-Playing Games (MMORPG). Games from this category allow users to establish and maintain social ties, by interacting and collaborating with strangers (Zhong, 2011). They support large numbers of players simultaneously in the same server (Islam et al., 2016). These games often maintain a rather stable number of players over time, as they sustain large audiences interested in the game who finance it so that they can continue to play (A. M. M. Santos et al., 2017). This is a critical aspect for our choice, as we focus on studying sentiment changes over time. Thus, avoiding videogames that lose interest shortly after release, and allowing us to increase the timespan of included reviews. In the next section we will explore the existent methods to extract sentiment from opinion text such as user reviews.

2.2. SENTIMENT ANALYSIS

Sentiment analysis is the computational study of people's opinions, attitudes and emotions toward an entity (Medhat et al., 2014). The target of sentiment analysis is to find opinions, identify the sentiments they express, and then classify their polarity. Sentiment analysis has been used in several applications including analysis of the repercussions of events in social networks, analysis of opinions about products and services, or simply to better understand aspects of social communication in social networks (Gonçalves et al., 2013).

More recently, sentiment analysis applications have been made in the context of COVID-19. Part of the studies consist on analysing social media messages, such as tweets, related to the pandemic, in order to detect mood changes (Nemes & Kiss, 2020). Results display, for example, mood deterioration correlated with lockdown announcements (Kruspe et al., 2020). Also using online discussions data, efforts were made into finding out the main concerns of people during the start of COVID-19 outbreak, highlighting the uncertainty around health and family, as well as origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection (Abd-Alrazaq et al., 2020; Li et al., 2020). Another study queried news articles from February 2020, to find that there were more negative articles than positive articles, as well as their subject (Hamzah et al., 2020). By analysing user reviews, the impact of the pandemic on user experience with online education platforms in China was assessed, identifying a change in concerns displayed before and after the pandemic outbreak (Chen et al., 2020).

Regarding sentiment analysis applications to videogames, past works include the analysis of in-game player chat messaging (Thompson et al., 2017), assorted game related tweets (Islam et al., 2016) and written reviews, which is the target of this study. A previous study on the differences between expert and amateur videogame reviews, concludes that amateur reviews are often more polarized, emotionally charged and not only written right after the videogame release (T. Santos et al., 2019). Many studies have been conducted on videogame amateur reviews in order to extract sentiment. A previous work from 2019 includes a review on the existent studies by then (Rajapakshe, 2019). This review detects two main types of works. Binary classification ones which evaluate if the sentiment displayed is positive or negative and compare it with whether the user recommended or not the game. Aspect-based classification ones with focus on separately studying feedback on certain features.

Among the existent sentiment analysis techniques, two main categories can be detected according to previous surveys (Medhat et al., 2014; Pradhan et al., 2016), Machine Learning (ML) and Lexicon-based, as seen in Figure 3.

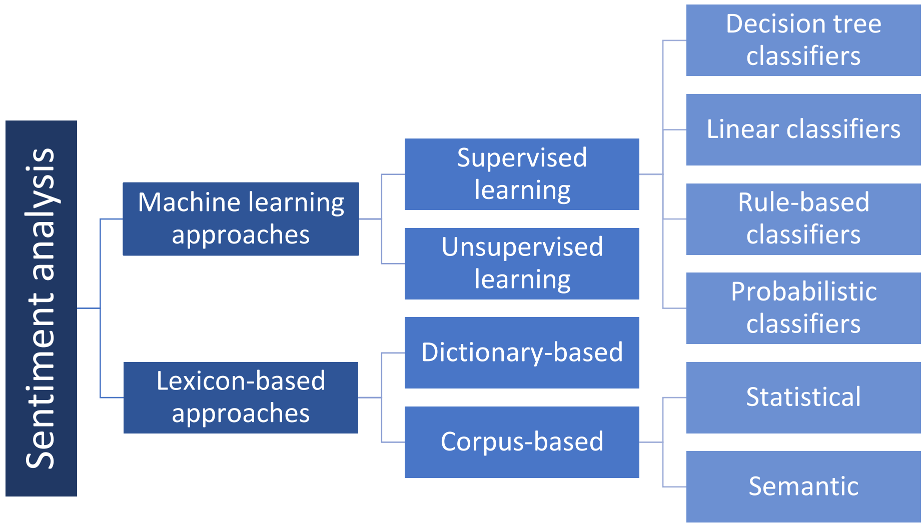

Figure 3 – Sentiment analysis approaches

2.2.1. Machine Learning approaches

These approaches can categorize themselves as supervised learning or unsupervised learning. The first category contains the vast majority of ML applications on sentiment analysis and we now focus on it. Supervised learning methods require the definition of two sets in the data, one for training and another for testing, to validate performance of the algorithm. These algorithms are mainly used on text classification problems, to predict one of the available classes for each record in the dataset. As example, previous work built a decision tree classifier with an accuracy of 0.75, to classify Steam user reviews as positive or negative, according to the sentiment in the text (Zuo, 2018). Another work on this purpose, besides using text variables also included available numerical variables such as "The number of hours a player plays a game before posting a review", obtaining their best results while using Support Vector Machines (SVM) and Term-Frequency and Inverse-Document-Frequency (TF-IDF) (Bais et al., 2017). A similar study used Artificial Neural Network and Classification and Regression Tree instead, obtaining better accuracy with the latter (Kang et al., 2017).

2.2.2. Lexicon-based approaches

Such approaches return the sentiment score for the target text, calculated via aggregation of the scores for every word (Gupta & Agrawal, 2020). For this purpose, pre-prepared Sentiment Lexicon is required, containing for every word in it defined, its semantic orientation, which can be positive, negative or neutral. Included in the Lexicon Based approaches category, there are the dictionary based, and the corpus-based methods:

- Dictionary based ones start by collecting a small set of words from the text, followed by dictionary expansion using common online dictionaries such as WordNet. Once this process is completed, error detection can be manually done. As an example, a previous approach extended a dictionary of lexicon-based sentiment extractor, to analyze in-game messages from the videogame StarCraft 2 (Thompson et al., 2017). This work displayed how Lexicon-based approaches can turn out as a useful and portable sentiment extraction method.
- Corpus-based allow the understanding of the context where relevant words for sentiment analysis are used in. There are two main methods for this approach. The statistical approach focuses on word occurrences. Here, polarity is defined as positive, negative, or neutral, according to the frequency the word displays in positive or negative text. The semantic approach works by giving sentiment values to words, while also considering semantically similar words, valuing them accordingly. This can be done by using a pre-defined corpus, to expand the initial set of synonyms and antonyms to the original word.

2.2.3. Chosen sentiment analysis methods

Considering the problem we have at hands and our goal to study sentiment changes over time, we will be going forward with two NLP Python libraries, TextBlob (Loria, 2018) and Valence Aware Dictionary and Sentiment Reasoner (VADER) (Hutto & Gilbert, 2014). Both algorithms are reported to work well on text found online, such as user-reviews (Bonta & Janardhan, 2019). The end product is also a reason for our choice, as the two methods return an overall sentiment polarity value, which we can later analyze in timelines.

As lexicon-based approaches, both libraries follow the process defined in Figure 4. This method begins with the tokenization of the gathered text, where the text is split into words. Follows the Part-Of-Speech (POS) Tagging part, consisting of converting the list of words into a list of tuples, including for each word, whether it is a verb, noun, or other part-of-speech. Each tuple is then assigned a sentiment score from the lexicon included in the package. Lastly, the output results are calculated from the scores of every word and normalized.

Figure 4 – Sentiment analysis process of the chosen approaches

TextBlob is a Python library for processing textual data. Its sentiment analysis feature uses lexicon resources from Natural Language ToolKit (NLTK) which is an open source NLP Python package. It

Returns two variables: sentiment polarity and subjectivity. The polarity score is a float within the range [-1.0, 1.0] where 1 is positive sentiment and -1 is negative sentiment, representing the calculated sentiment score of a text. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. It is used to measure whether the text is mostly based on personal opinion (High subjectivity) or in factual information (Low subjectivity).

Cleant review: "I love this game, it is fun to play with friends. Servers can be a problem."

TextBlob sentiment polarity: 0.133

TextBlob sentiment subjectivity: 0.400

VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, also working well on texts from other domains. It returns sentiment polarity indices of the target text in the form of three scores (Positive, Negative and Neutral) which represent the proportion of text that falls in these categories. It also returns the Compound metric, which is the calculated sum of all the lexicon ratings that have been normalized between -1 (most extreme negative) and +1 (most extreme positive).

Cleant review: "I love this game, it is fun to play with friends. Servers can be a problem."

VADER sentiment polarity: {'neg': 0.109, 'neu': 0.364, 'pos': 0.526, 'compound': 0.8834}

Once we have applied the two methods to the collected reviews from Steam and sentiment polarity values are saved, we will be ready to move on to the next step.

2.3. TESTS FOR STRUCTURAL BREAKS IN TIME-SERIES

Keeping in mind the main goal, a final step is required in order to identify the main changes over time in user reviews number and sentiment, as well as in the number of active players. This step consists in studying in timelines the data previously obtained and testing for the existence of structural breaks. Structural break points of time series are defined as points in a regression, that when used to split a time series, generate two parts with heavily dissimilar regression parameters. While the amount of work on this topic is vast, applications are mostly focused in the financial and economic fields. Past works include forecasting stock return volatility, demonstrating the challenges presented by structural breaks (Rapach et al., 2008). Another work regards the existence of structural breaks on sentiment analysis output data. This by collecting social media data from Twitter and StockTwitts, extracting sentiment series and proving the importance of detecting structural breaks, when using online investor sentiment series (Ballinari & Behrendt, 2020). It is, however, noticeable, the lack of research on structural breaks for videogame or user review related data, an issue that our research is expected to help mitigate.

When considering structural break detection tests, methods can date far back to 1960, with the Chow test. This widely known test operates by dividing the sample data into two subperiods. It estimates the parameters for each subperiod and uses a classic F statistic to validate the equality between the two sets of parameters (Hansen, 2001). The Chow test is held by its main limitation, requiring the assumption of the breakdate before performing the test. While popular for many years, more advanced alternatives were developed over time.

A popular example is the 'supremum' test of Andrews (Andrews, 1993). This test defines a supremum version of the F statistic. The sup-F uses a standard distribution under the null hypothesis, hence moving the change point parameter to the alternative hypothesis. As a result, it provides a work-around the Chow test referred limitation, enabling the test for an unknown change point.

A further approach by Bai and Perron (Bai & Perron, 2003) regards the issue of testing for structural changes on data and errors under very general conditions. Most importantly, it allows testing for multiple structural breaks without the need to previously specify the break dates. This was achieved via a dynamic-programming-based algorithm, used to obtain global minimizers of the sum of squared residuals.

Instead of directly applying one of the previous methods, we chose to use an R package, Strucchange (Zeileis et al., 2002), to define and conduct our structural break tests. This package provides the tools to apply those methods through a set of defined functions. Considering the data we are studying, and the uncertainty regarding the existence of one or more structural changes, we intend to test for the existence of structural breaks. The tests will be created, taking into account previous applications of this package (Zeileis et al., 2003; Zeileis & Kleiber, 2005). Here, we regard the standard linear regression model

$$y_t = x_t^T \beta_i + \varepsilon_t \quad (t = 1, \ldots, T), \tag{1}$$

where i is the time and y_i is the observation of the variable being tested, x_t is a $k \times 1$ vector of regressors, having the first component usually equal to unity, and β_i is the $k \times 1$ vector of regression coefficients which may vary over time.

In this work, we intend to test the hypothesis that the regression coefficients (β_i) remain constant

$$H_0: \quad \beta_i = \beta \quad (i = 1, \ldots., n), \tag{2}$$

versus the alternative (H_1) that at least one of the coefficients varies over times. The rejection of the null hypothesis would then imply the existence of structural breaks in the models.

From the collected data we selected the following four variables for testing: Monthly total amount of reviews; Monthly mean of sentiment polarity scores obtained with TextBlob; Monthly mean of the compound metric of sentiment polarity indices obtained with VADER; Monthly total of players. The timespan of the dataset ranges from the beginning of January 2018 to the end of April 2021. Hence, including data from a total of 40 months. The data at this point collected and pre-processed will be extracted from Python into R using CSV files. From here, a set of tests will be conducted for each of the variables, to investigate and identify the existence of structural changes. In our scenario, considering our uncertainty regarding the number of breaks in the data, we will leave the maximum number of breaks parameter set to be $M = 5$ and trimming parameter of 15% of the length of our data. This set consists of the following tests:

Test 1: The first test will focus on detecting abrupt structural changes in the mean of the series. We will conduct this test by fitting the following model:

$$y_t = \beta_0 + \varepsilon_t \tag{3}$$

Test 2: On the second test we study the existence of broken trend (t denotes a deterministic linear trend). As follows:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{4}$$

Test 3: The third test detects structural breaks on the parameters of an auto-regressive model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \tag{5}$$

Once conducted the tests on the four selected variables, the outputted breakpoints will be stored in a table containing the results for all games, for further analysis. Confidence intervals for each resultant break date, as well as the p-value for each test will also be collected.

3. RESULTS & DISCUSSION

3.1. DATA COLLECTION

Before extracting user review data, we researched on the existent methods to obtain such data from Steam and found various ways to do it. A previous approach on sentiment classification, extracted reviews from Steam using a sample Javascript code that identifies the useful HTML elements from the webpage and extracts those useful features and texts into written JSON file (Bais et al., 2017). Another work proposes a method that includes extracting reviews using a crawler that performs HTTP REST requests to the Steamworks API. Followed by transforming review text through operations such as lowering case, removing URLs and replace internet slang with textual counterparts (Vieira & Brandão, 2019).

Our method choice, however, fell on an existent method, that uses a Python API instead of creating a Web Scraping script (Zuo, 2018). The API we used is named "steamreviews", and all it needs in order to retrieve user reviews for a particular game is the correspondent game ID on Steam, having a few more optional parameters. The first step was to get IDs and names for all the target games and to save the IDs into a list. The second step was to use the reviews API with each ID to get all the reviews for each game. Regarding the API parameters, we decided to collect only reviews that are written in English, to make the sentiment analysis task more practical. For this reason, besides requesting only reviews written after the start of 2018, we specified the language parameter as English only.

User review extraction was conducted in a Python Jupyter Notebook. Python functions were developed in order to download the reviews using the API, which generates one JSON file per game. Such files contained the large variety of features present in each Steam user review, from which we initially selected the features displayed in Table 1, before transforming data into a CSV data table format, to be used in the analysis. We opted to select the date of last update on the review, instead of the date of writing, due to users being able to modify their review over time, sometimes reflecting an update which improved or decreased their experience of the game.

Table 1 – Steam review features

Feature	Description
Author Playtime total	Total playtime in hours by the player on the reviewed game.
Review text	Actual text of the review.
Date of last update	Date when the review was last updated.
Recommendation	Whether the user marks the game as Recommended or Not Recommended.
Votes up	Number of votes up given by other users to the review.
Votes funny	Number of votes funny given by other users to the review.

For our study, we considered it would be beneficial to include active players numbers in our analysis dataset. This data may help understanding popularity changes of the target videogames and even videogames in general. Such data may not be obtained in a simple way directly from Steam. However, there are free third-party tools that collect and process this type of data. Steam Spy is an example, which continuously collects data from Steam, including historical values of the number of owners and players for each game (Lin et al., 2019). Another useful third-party application is Steam Database. Same case as Steam Spy, Steam DB provides detailed insights of raw data obtained from Steam, in a more comprehensive way than the official API (Zuo, 2018).

Steam DB turned out to be the best option to obtain Steam player data as it allows downloads in the form of CSV or XLS files, containing the features displayed in Table 2. The downloaded files include not only the daily player count since the game release, but also the daily amount of Twitch viewers, watching streams of the specific game in particular. Twitch is a large live streaming platform with main focus on videogames (Hilvert-Bruce et al., 2018), and its data can be particularly useful in our study. The files also contained a flag variable representing if there was a major update or sale on the correspondent day. However, it turned out to be incomplete or fully empty for most games, reason why we will not be using its data.

Table 2 – Data collected from SteamDB

Feature	Description
DateTime	Day of the year.
Players	Daily total number of Steam players.
Twitch	Daily total number of Twitch viewers.

3.2. DATA PRE-PROCESSING

The pre-processing phase is a crucial step when building effective sentiment analysis systems for text mining (Angiani et al., 2016). Online texts such as reviews often contain lots of noise, irrelevant data and even sarcasm. Although our chosen sentiment analysis algorithms support online text with close to none pre-preparation, we will clean the data regardless, in order to reduce the size of the dataset and improve processing times. For this we used our own set of techniques, which includes the following pre-processing steps:

- Removing links: Links may not contain sentiment information, so we removed them from the dataset. A regular expression is used to identify strings that start with "https://", removing them afterwards.
- Removing special characters and digits: Non-letter characters represent unnecessary crowd on data, since they are not useful for the sentiment analysis algorithm. They were removed by keeping only the upper and lower cases of English letters.
- Lower case every letter: By this step, every non letter character was removed. To reduce the size of words, we transformed every letter in reviews into lower case.
- Removing stop words: Stop words are words which are filtered out before or after processing of natural language data. Examples of stop words include "a", "and", "the", "how" and "or". While these words do not have sentiment values, they highly increase the weight of the dataset. To remove them we used the NLTK Python package.

Before:

"Best Sub based MMO on the market. Amazing story. Each expansion has only gotten better. Well worth the money and time!"

After:

"best sub based mmo market amazing story expansion gotten better well worth money time"

After the pre-processing steps were conducted, the total amount of words from all reviews was nearly reduced to half (20.1 million words to 10.4 million). The most common words in our review dataset can be seen in Table 3. It is noticeable the predominance of words with positive connotation amongst the list, such as "good", "fun", "great", "best", etc. Furthermore, the presence of the words "people" and "friends" confirm the social component of the target videogames, known for allowing players to meet new people and play with their friends online.

Table 3 – Most common words

Word	Count	Word	Count	Word	Count
game	530,095	one	54,795	people	40,965
good	124,460	hours	51,801	played	40,086
fun	123,349	even	48934	love	40,082
play	121,057	playing	46,635	still	38,969
like	101,756	best	46,236	want	36,439
get	91,681	free	45,195	lot	35,985
time	75,856	new	45,021	grind	33,775
great	67,829	much	44,362	players	33,028
dont	61,558	games	42,755	friends	32,761
really	56,830	would	40,973	make	32,094

3.3. COLLECTED DATA

The data obtained corresponds to user-reviews, scores and player count on 20 of the most popular MMORPGs on Steam (ARK Survival Evolved, Black Desert Online, Crossout, DC Universe Online, EVE Online, Final Fantasy XIV, Fishing Planet, Neverwinter, Path Of Exile, Smite, Realm of the Mad God Exalt, Spiral Knights, The Elder Scrolls Online, Tree of Savior (Eng. Version), Trove, Warframe, VRChat, War Thunder, World of Tanks Blitz, World of Warships, Trove). As we aimed to have the possibility to compare trends and identify patterns before and after the COVID-19 outbreak, we opted to include reviews written between January 2018 and April 2021 in our study. For this reason, we only selected games that were released before January 2018.

We now take an overall look at the data obtained. Main data insights extracted, are displayed by videogame, in

Table 4. The final dataset containing data from the 20 selected games, obtained after data pre-processing steps consists of nearly 600.000 reviews, with a total of 10.4 million significant words. Having the reviews of all games together, in 83.2% of them, the user recommends the game. If calculated separately for each game, the average recommendation rate between the 20 games is 76.0%. This suggests that the most popular games from the 20, have a higher positive recommendation rate by the players who wrote the reviews. Our dataset was posteriorly grouped by date of last update on reviews, in months, to study possible changes over time. Each record corresponding to a month between January 2018 and April 2021 and containing all the necessary features we collected.

Table 4 – Descriptive statistics of the 20 games

Game Name	Total Reviews	Recommended	Avg Word Length	Avg Playtime (hrs)	Avg Monthly Players	Avg Monthly Twitch Viewers
ARK	106,265	85%	15	724	1,808,632	278,680
Black Desert Online	22,596	73%	26	1494	537,776	No data found
Crossout	10,616	74%	23	281	156,747	9,081
DC Universe Online	3,052	75%	19	231	21,617	4,306
EVE Online	8,242	76%	26	664	137,542	38,185
FINAL FANTASY XIV Online	18,884	89%	24	1541	498,480	No data found
Fishing Planet	5,917	85%	13	164	66,532	10,163
Neverwinter	5,454	75%	23	309	91,479	8,792
Path of Exile	59,834	93%	16	1056	1,222,810	628,561
Realm of the Mad God Exalt	10,576	84%	12	506	58,582	6,023
Smite	28,719	78%	16	672	474,766	263,214
Spiral Knights	3,451	83%	19	213	10,459	853
The Elder Scrolls Online	38,228	82%	25	696	665,370	291,512
Tree of Savior (Eng. Version)	2,883	66%	26	694	66,932	2,889
Trove	14,043	81%	11	245	111,673	6,836
VRChat	55,936	91%	11	407	317,097	228,882
War Thunder	61,884	79%	18	610	694,653	49,709
Warframe	108,298	93%	18	898	1,988,001	414,146
World of Tanks Blitz	9,130	79%	14	388	646,462	4,255
World of Warships	21,890	84%	19	449	294,075	183,936

3.4. Structural Break Tests Results

Figure 5 provides a glance of the amount of reviews written over time. The blue part of the histogram bars represents the reviews where the game is recommended by the user, while the red part represents the ones where the game is not recommended. While the punctual rise in the amount of reviews dating from November 2018 and June/July 2019 may be justified by sales on the Steam store, the sudden increase dating from November 2019 is a different scenario. Besides the Black Friday Sales, the Big Library Update had its beta release in September 2019 and full release at the very end of October 2019. As it introduced new ways to prompt the user to write a review, the number of monthly written reviews suffered a very large boost.

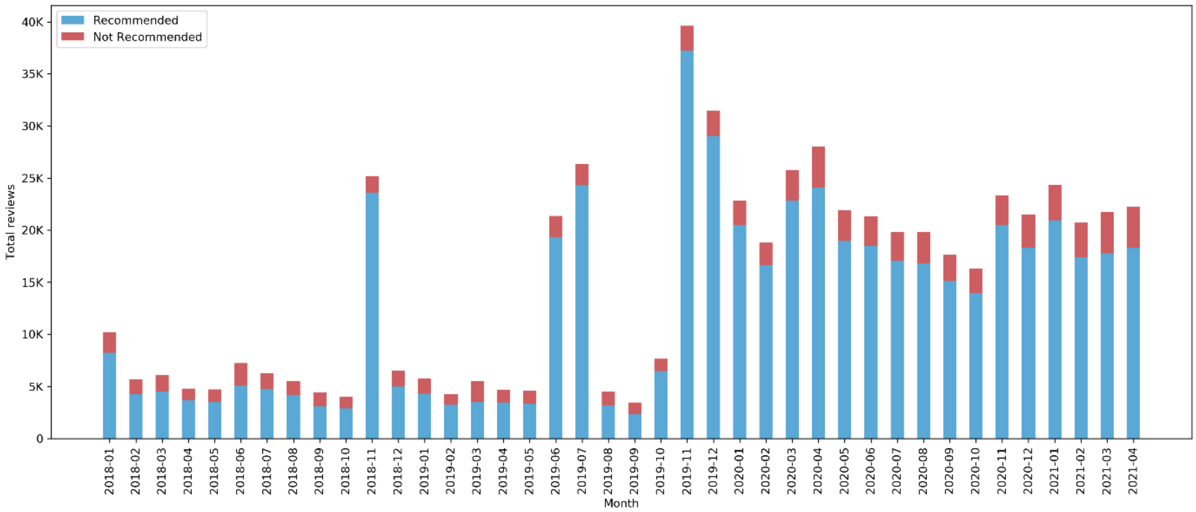

Figure 5 – Total of reviews since 2018 per month

Regarding Figure 6, this figure shows the percentage of positive and negative reviews since 2018 per month, as well as the sentiment polarity scores obtained. Following the same logic as in Figure 5, the blue part of the histogram bars represents the reviews where the game is recommended by the user, while the red part represents the ones where the game is not recommended. The black and the blue lines represents the sentiment polarity scores obtained, being the black line the polarity value from TextBlob and the blue line the compound metric from VADER. Comparing the percentage of positive reviews with the sentiment polarity scores, it is generally noticeable how the sentiment polarity values tend to be higher, the higher the percentage of positive reviews during the same month, as expected. Furthermore, when comparing Figure 5 with Figure 6, it is perceptible that in months where the number of reviews rose, the percentage of positive reviews compared to negative ones also seemed to increase.

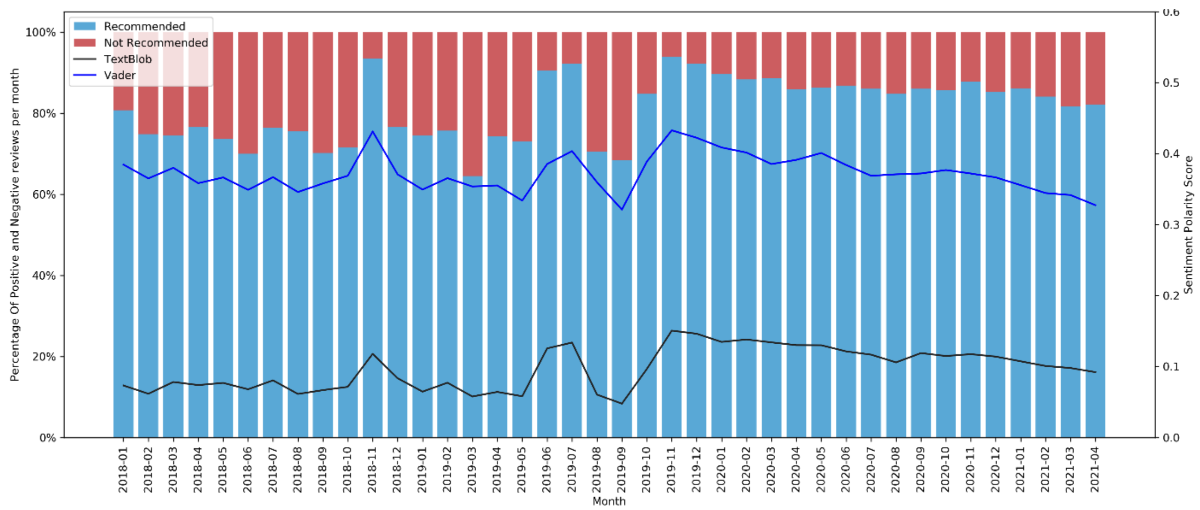

Figure 6 – Percentage of positive and negative reviews and sentiment polarity scores since 2018 per month

Now we look at the active players numbers, which is the red line from Figure 7. This data represents the monthly sum of daily active players for the selected videogames. By observing its evolution over time, we can easily notice the large increase in players that happened by March 2020. Such event corresponds to the COVID-19 pandemic global outbreak date, closely followed by the quarantine measures, which forced people to stay at home. On the months that followed, player numbers lowered as the measures softened but were still high when compared to the two previous years, as people were still advised to stay at home. Figure 7 also includes represented as a blue line, the monthly sum of daily Twitch viewers for 18 out of our 20 target games, as two of them did not have this data available. Despite not showing the same of impact of the pandemic outbreak as in the active player numbers, the amount of Twitch viewers was still fairly high over the year of 2020, when compared with the numbers from 2018 and 2019.

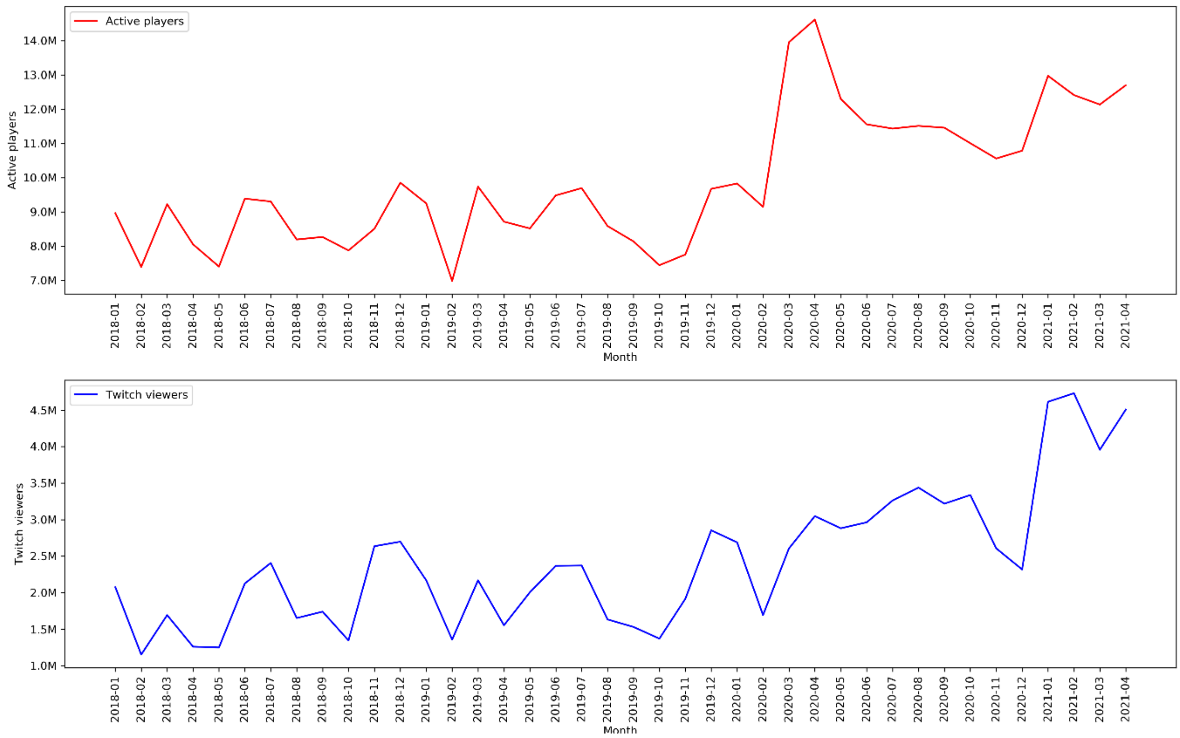

Figure 7 – Active players and Twitch viewers

In an effort to obtain a clear answer regarding the possible impact of the pandemic on user reviews and player numbers, the previously explained set of structural change tests was conducted. In Figure 8, are displayed the results from the application of the 3 tests created, to the 4 selected variables, with all data from the 20 games together. While the black lines in Figure 8 display the assessed variables, the red dotted lines represent the structural breaks found. Regarding the amount of reviews, the first two tests detected a structural change in October 2019. Precisely by the time Big Library Update went live, which caused an increase in the amount of reviews being written by users. This structural break was also detected by three of the six tests conducted to the sentiment polarity scores variables plus another test (VADER & Test 1) that detected a break just one month before October 2019. Moreover, the first test detected a change in the sentiment polarity scores from both algorithms on the first half of 2020, during the months that shortly followed the worldwide COVID-19 pandemic outbreak in March 2020. Regarding the number of players, the first two tests detected a break by February 2020, just before the global COVID-19 spread.

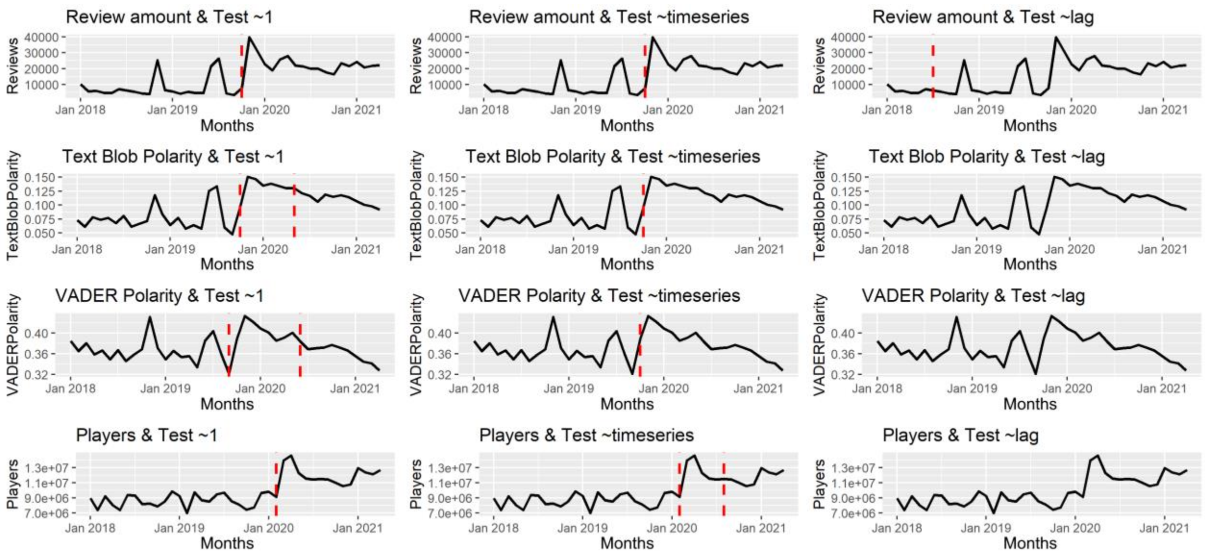

Figure 8 – Structural break tests results with data from all 20 games

Table 5 provides additional information on the structural break test results displayed in Figure 8. For each test the correspondent p-value is displayed, as well as the confidence interval for each break detected, representing the margin of error. Considering the applications of Test 2, and applications of Test 1 apart from VADER Polarity, we can reject the null hypothesis for those tests. This way, confirming the existence of structural breaks on the tested variables.

Table 5 – Structural break tests result with data from all 20 games

Test	Break Date	p-value	95% Confidence Interval
Review amount & Test 1	Oct-2019	<0.01	Aug-2019 – Jan-2020
Review amount & Test 2	Oct-2019	<0.01	Aug-2019 – Nov-2019
Review amount & Test 3	Jul-2018	<0.01	None (Start date included in trimmed months)
Text Blob Polarity & Test 1	Oct-2019, May-2020	<0.01	Sep-2019 – Dec-2019, Mar-2020 – Jun-2020
Text Blob Polarity & Test 2	Oct-2019	<0.01	Sep-2019 – Nov-2019
Text Blob Polarity & Test 3	None	<0.01	None
VADER Polarity & Test 1	Sep-2019, Jun-2020	0.1	Jul-2019 – Fev-2020, Apr-2020 – Aug-2020
VADER Polarity & Test 2	Oct-2019	<0.01	Sep-2019 – Nov-2019
VADER Polarity & Test 3	None	<0.01	None
Player count & Test 1	Feb-2020	<0.01	Dec-2019 – Mar-2020

Player count & Test 2	Feb-2020, Aug-2020	<0.01	Jan-2020 – Mar-2020, Jul-2020 – Sep-2020
Player count & Test 3	None	<0.01	None

Considering that the 5 most popular from the 20 selected games generated 66% of the total number of reviews collected, the tests for structural changes have also been done per game individually. The results obtained can be seen in the figures displayed in Appendix A - Regression and structural break test results per game. The application of the 3 break tests to the 4 selected variables of the 20 selected games resulted in a total of 240 test combinations. From this, 272 breaks were globally detected, resulting in a mean of slightly over 1 break found per test. 111 breaks come from the applications of structural break test 1, 114 from test 2, with test 3 detecting only 47 breaks, less than half when compared with the other 2 tests. Using data from the tests, Figure 9 displays the total of breaks detected by month for each study variable. Here, the red line represents the amount of written reviews, and displays the highest concentration of breaks detected (34) between all variables, on October 2019 (Steam Big Library Update). **T**he TextBlob sentiment polarity, along with the VADER sentiment polarity are associated with the blue and green lines respectively. These variables also show the highest number of breaks by September & October 2019 due to the same reason as the review amount. The player numbers variable, corresponding to the yellow line however, had the major number of breaks detected by the time of the pandemic outbreak.

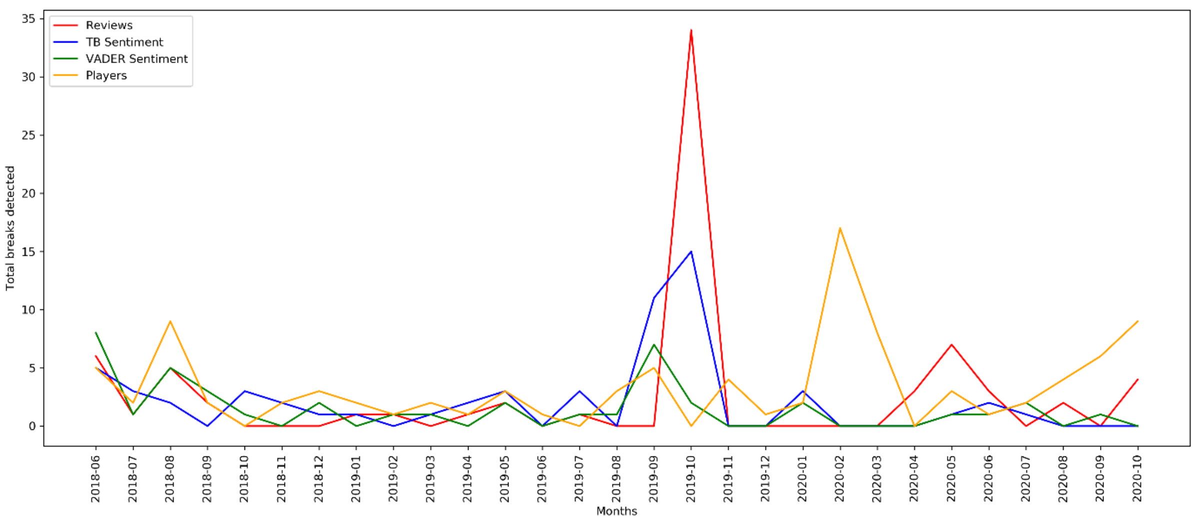

Figure 9 – Monthly distribution of all breaks detected (3 tests x 20 games) per variable

4. CONCLUSION

The goal of this study was to assess the impact caused by large-scale events such as the COVID-19 pandemic and major updates on videogames and their platform. User reviews and active player data from Steam was extracted. Two sentiment analysis methods, TextBlob and VADER were chosen and applied to Steam user reviews. Ultimately, three different structural break tests were prepared and applied to four selected variables, using Strucchange R package. After a careful analysis of the obtained results, we lack proof to assume that the COVID-19 pandemic outbreak did in fact affect the sentiment displayed towards videogames, as the impact on Steam user reviews was not significant enough. Nonetheless, it was possible to identify as structural breaks, the change caused by the pandemic in the number of active players, which greatly increased in the months that followed the outbreak. During our research we discovered the Steam Big Library Update and how it turned out to have a large effect in Steam reviews. As the number of written reviews notoriously increased after this change, the percentage of positive reviews rose by 25% in the space of two months. Therefore, the sentiment contained in user reviews was also affected, and our tests detected structural breaks by the time of the update, validating its large impact.

Detecting structural breaks in a time series, is a valid way to assess the existence of critical changes in the data. Our approach applies structural break testing in a whole new context, providing a new method for videogame developers and reviewing platform managers to assess game and reviewing related data, such as the sentiment expressed by players. This way, understanding the repercussions of launched updates and/or the impact of outside events, as well as obtaining a clearer view on how to act accordingly.

Future work could include applying this method to other feedback sources, such as social media or in-game player chat, to study updates on a specific videogame or platform. Another option would be joining data from different sources in a single study, which may provide better results when assessing large-scale worldwide events, mitigating the impact of platform related updates on data. More classical approaches for sentiment analysis can also be put into practice and different multivariate structural tests can be created using different formulas to fit data.

5. REFERENCES

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, *22*(4), e19016.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 821–856.

Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., & Manicardi, S. (2016). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. *KDWeb*.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, *13*(3), 315–352.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Bais, R., Odek, P., & Ou, S. (2017). *Sentiment Classification on Steam Reviews*. Stanford University.

Ballinari, D., & Behrendt, S. (2020). Structural breaks in online investor sentiment: A note on the nonstationarity of financial chatter. *Finance Research Letters*, *35*, 101479.

Bonenfant, M., Deslauriers, P., Dumont, A., Jobin, A., & Pelletier-Gagnon, J. (2020). "Finally! My first shiny!": Assessing Text Mining Tools and Methodologies for the Study of the Pokémon Sword and Shield Twitter Online Community. *CSDH-SCHN 2020*.

Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1–6.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*.

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment*, 138882.

Chen, T., Peng, L., Jing, B., Wu, C., Yang, J., & Cong, G. (2020). The impact of the COVID-19 pandemic on user experience with online education platforms in China. *Sustainability*, *12*(18), 7329.

Farboodi, M., Jarosch, G., & Shimer, R. (2020). *Internal and external effects of social distancing in a pandemic*. National Bureau of Economic Research.

Farooq, A., Laato, S., & Islam, A. N. (2020). Impact of online information on self-isolation intention during the COVID-19 pandemic: Cross-sectional study. *Journal of Medical Internet Research*, *22*(5), e19128.

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. *Proceedings of the First ACM Conference on Online Social Networks*, 27–38.

González-Piñero, M. (2017). Redefining the value chain of the video games industry. *Memory*, *36*(4), 75–90.

Gupta, N., & Agrawal, R. (2020). Application and techniques of opinion mining. In *Hybrid Computational Intelligence* (pp. 1–23). Elsevier.

Hagen, M., Potthast, M., Büchner, M., & Stein, B. (2015). Webis: An ensemble for twitter sentiment detection. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 582–589.

Hamzah, F. B., Lau, C., Nazri, H., Ligot, D., Lee, G., Tan, C., & others. (2020). CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, *1*(32).

Hansen, B. E. (2001). The New Econometrics of Structural Change: Dating Breaks in U.S. Labour Productivity. *Journal of Economic Perspectives*, *15*(4), 117–128.

Hilvert-Bruce, Z., Neill, J. T., Sjöblom, M., & Hamari, J. (2018). Social motivations of live-streaming viewer engagement on Twitch. *Computers in Human Behavior*, *84*, 58–67.

Hu, G., Bhargava, P., Fuhrmann, S., Ellinger, S., & Spasojevic, N. (2017). Analyzing users' sentiment towards popular consumer industries and brands on Twitter. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 381–388.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1).

Ingram, J., Maciejewski, G., & Hand, C. J. (2020). Changes in diet, sleep, and physical activity are associated with differences in negative mood during COVID-19 lockdown. *Frontiers in Psychology*, *11*, 2328.

Islam, S. T., Al-Sajid, A., & Hussein, M. R. (2016). Design and simulation of a novel classification framework for separating sentiment from assorted game related tweets. *2016 International Workshop on Computational Intelligence (IWCI)*, 58–63.

Kang, H.-N., Yong, H.-R., & Hwang, H.-S. (2017). A Study of analyzing on online game reviews using a data mining approach: STEAM community data. *International Journal of Innovation, Management and Technology*, *8*(2), 90.

Kasper, P., Koncar, P., Santos, T., & Gütl, C. (2019). On the role of score, genre and text in helpfulness of video game reviews on metacritic. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 75–82.

Kruspe, A., Häberle, M., Kuhn, I., & Zhu, X. X. (2020). Cross-language sentiment analysis of European Twitter messages duringthe COVID-19 pandemic. *ArXiv Preprint ArXiv:2008.12172*.

Laato, S., Islam, A. N., & Laine, T. H. (2020). Did location-based games motivate players to socialize during COVID-19? *Telematics and Informatics*, *54*, 101458.

Laato, S., Laine, T. H., & Islam, A. (2020). Location-based games and the covid-19 pandemic: An analysis of responses from game developers and players. *Multimodal Technologies and Interaction*, *4*(2), 29.

Li, S., Wang, Y., Xue, J., Zhao, N., & Zhu, T. (2020). The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. *International Journal of Environmental Research and Public Health*, *17*(6), 2032.

Lin, D., Bezemer, C.-P., Zou, Y., & Hassan, A. E. (2019). An empirical study of game reviews on the Steam platform. *Empirical Software Engineering*, *24*(1), 170–207.

Loria, S. (2018). Textblob Documentation. *Release 0.15*, *2*.

Marston, H. R., & Kowert, R. (2020). What role can videogames play in the COVID-19 pandemic? *Emerald Open Research*, *2*(34), 34.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.

Nemes, L., & Kiss, A. (2020). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 1–15.

Newzoo. (2020). *Global Games Market to Generate $175.8 Billion in 2021; Despite a Slight Decline, the Market Is on Track to Surpass $200 Billion in 2023*. https://newzoo.com/insights/articles/global-games-market-to-generate-175-8-billion-in-2021-despite-a-slight-decline-the-market-is-on-track-to-surpass-200-billion-in-2023/

Pradhan, V. M., Vala, J., & Balani, P. (2016). A survey on sentiment analysis algorithms for opinion mining. *International Journal of Computer Applications*, *133*(9), 7–11.

Quader, N., Gani, M. O., Chaki, D., & Ali, M. H. (2017). A machine learning approach to predict movie box-office success. *2017 20th International Conference of Computer and Information Technology (ICCIT)*, 1–7.

Rajapakshe, U. (2019). Development Centric Player Feedback Analysis for Video Games: A Review. *2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, 190–194.

Rapach, D. E., Strauss, J. K., & Wohar, M. E. (2008). Forecasting stock return volatility in the presence of structural breaks. In *Forecasting in the presence of structural breaks and model uncertainty*. Emerald Group Publishing Limited.

Riva, G., Mantovani, F., & Wiederhold, B. K. (2020). Positive Technology and COVID-19. *Cyberpsychology, Behavior, and Social Networking*, *23*(9), 581–587.

Roy, A., Khan, M. H., & Chakraborty, S. (2018). *Analyzing users' sentiment towards video games based on reviews from microblog* [PhD Thesis]. BRAC University.

Ruseti, S., Sirbu, M.-D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2020). Comprehensive Exploration of Game Reviews Extraction and Opinion Mining Using NLP Techniques. *Fourth International Congress on Information and Communication Technology*, 323–331.

Santos, A. M. M., Franco, A., Maia, J. G. R., Gomes, F., & Castro, M. (2017). A methodology proposal for MMORPG content expansion analysis. *XVI Simpósio Brasileiro de Jogos e Entretenimento Digital*.

Santos, T., Lemmerich, F., Strohmaier, M., & Helic, D. (2019). What's in a Review: Discrepancies Between Expert and Amateur Reviews of Video Games on Metacritic. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–22.

Steam Database. (2020). https://steamdb.info/app/753/graphs/

Thompson, J. J., Leung, B. H., Blair, M. R., & Taboada, M. (2017). Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. *Knowledge-Based Systems*, *137*, 149–162.

Vieira, A., & Brandão, W. (2019). Evaluating Acceptance of Video Games using Convolutional Neural Networks for Sentiment Analysis of User Reviews. *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 273–274.

Zagal, J. P., Ladd, A., & Johnson, T. (2009). Characterizing and understanding game reviews. *Proceedings of the 4th International Conference on Foundations of Digital Games*, 215–222.

Zeileis, A., & Kleiber, C. (2005). Validating Multiple Structural Change Models: A Case Study. *Journal of Applied Econometrics*, *20*(5), 685–690.

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, *44*(1–2), 109–123.

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software, Articles*, *7*(2), 1–38.

Zhong, Z.-J. (2011). The effects of collective MMORPG (Massively Multiplayer Online Role-Playing Games) play on gamers' online and offline social capital. *Computers in Human Behavior*, *27*(6), 2352–2363.

Zuo, Z. (2018). *Sentiment analysis of steam review datasets using naive bayes and decision tree classifier*.

6. APPENDIX

6.1. APPENDIX A - REGRESSION AND STRUCTURAL BREAK TEST RESULTS PER GAME

Black lines display the assessed variables, while red lines represent the structural breaks found.

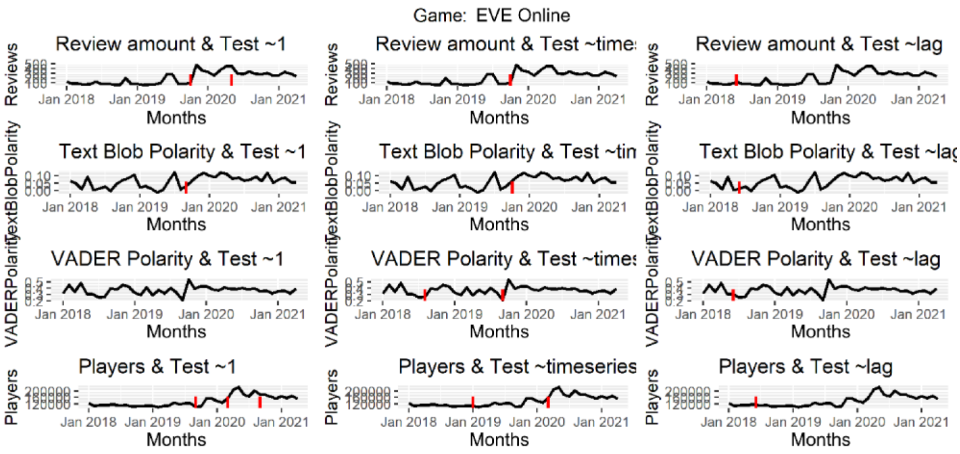

Figure 10 – Structural break test results: EVE Online

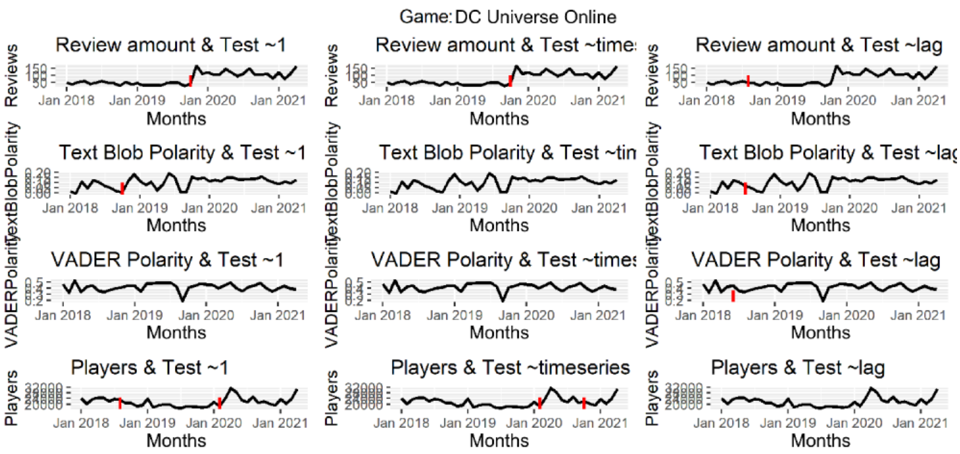

Figure 11 - Structural break test results: DC Universe Online

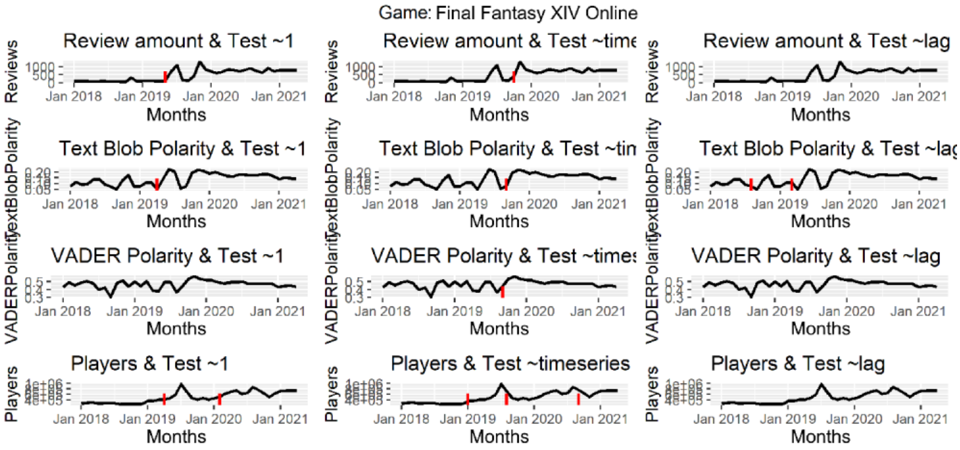

Figure 12 - Structural break test results: Final Fantasy XIV Online

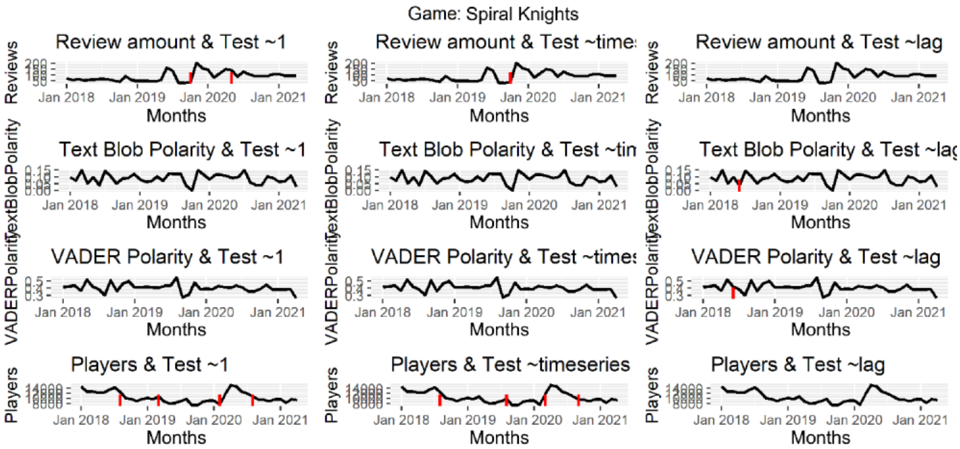

Figure 13 - Structural break test results: Spiral Knights

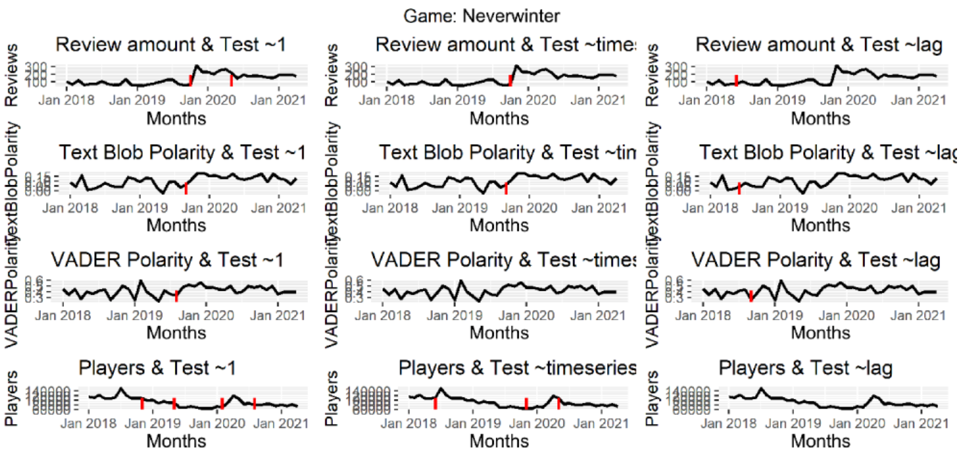

Figure 14 - Structural break test results: Neverwinter

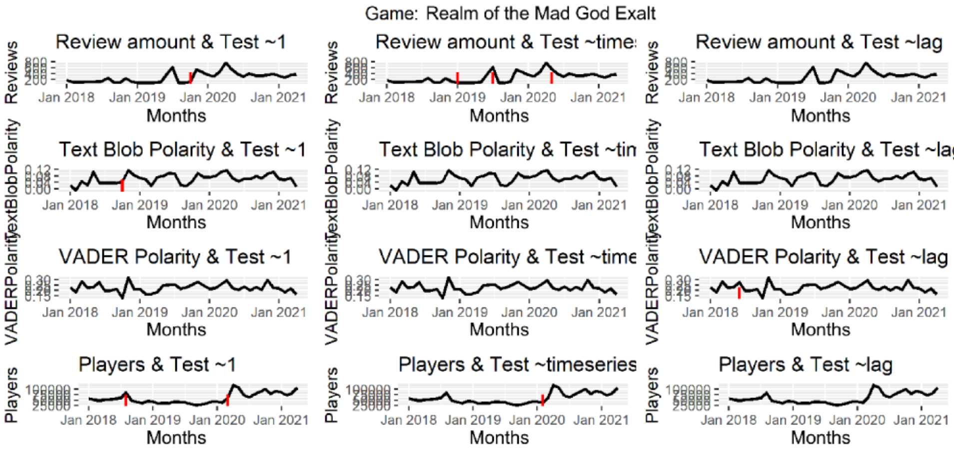

Figure 15 - Structural break test results: Realm of the Mad God Exalt

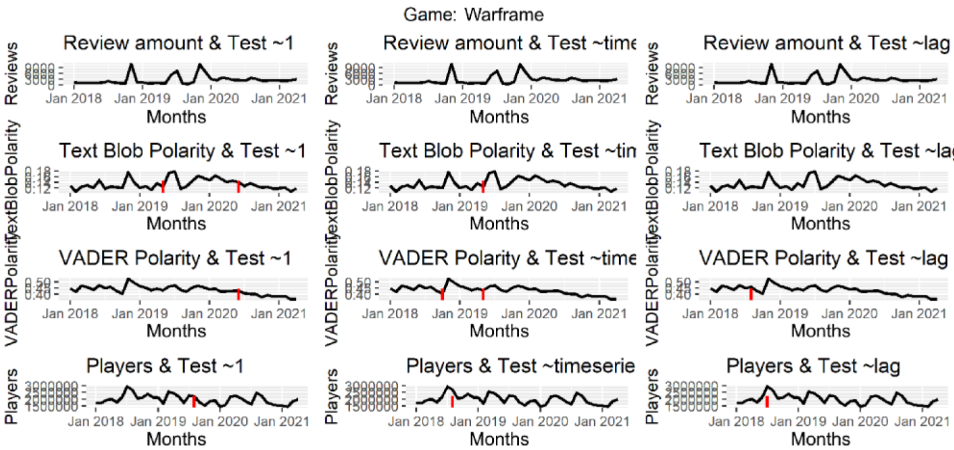

Figure 16 - Structural break test results: Warframe

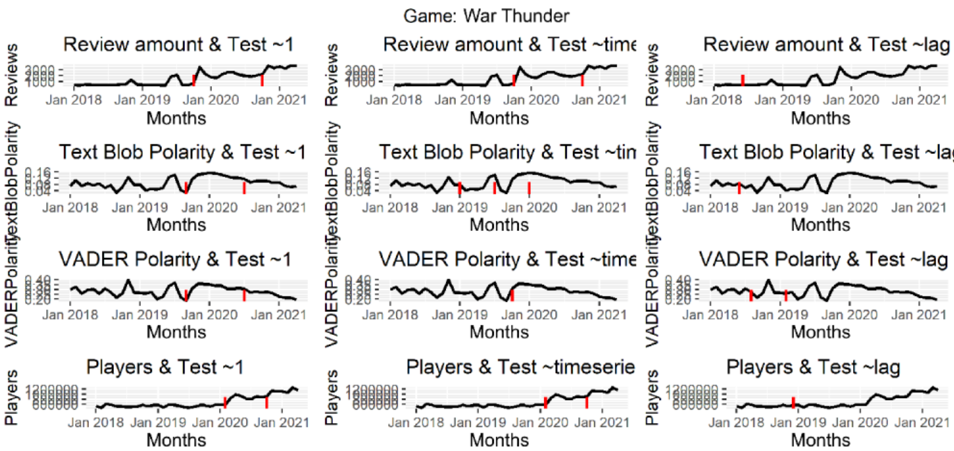

Figure 17 - Structural break test results: War Thunder

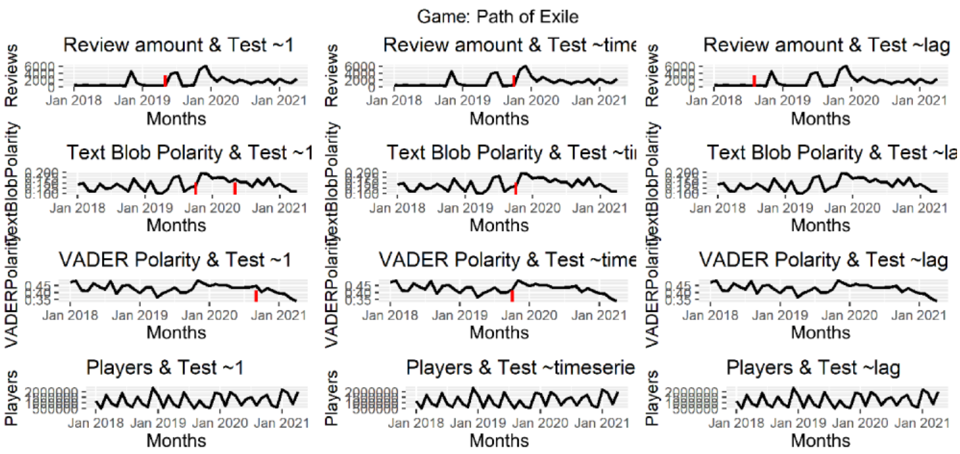

Figure 18 - Structural break test results: Path of Exile

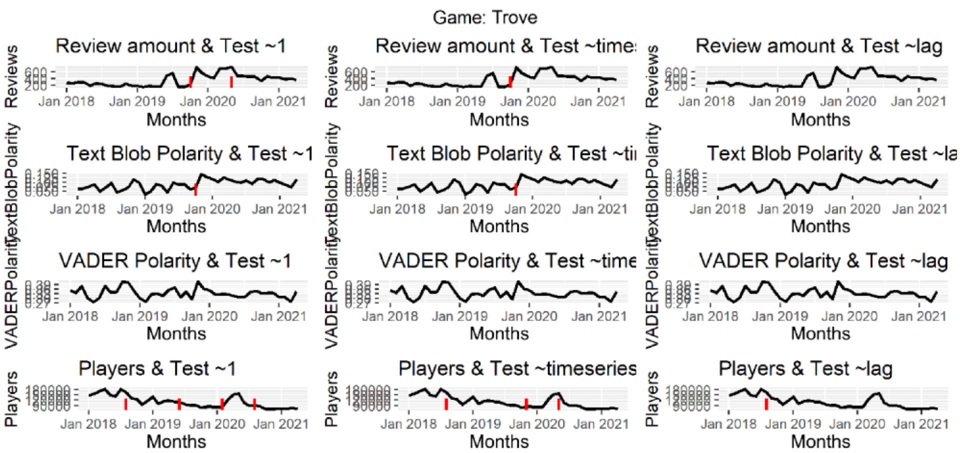

Figure 19 - Structural break test results: Trove

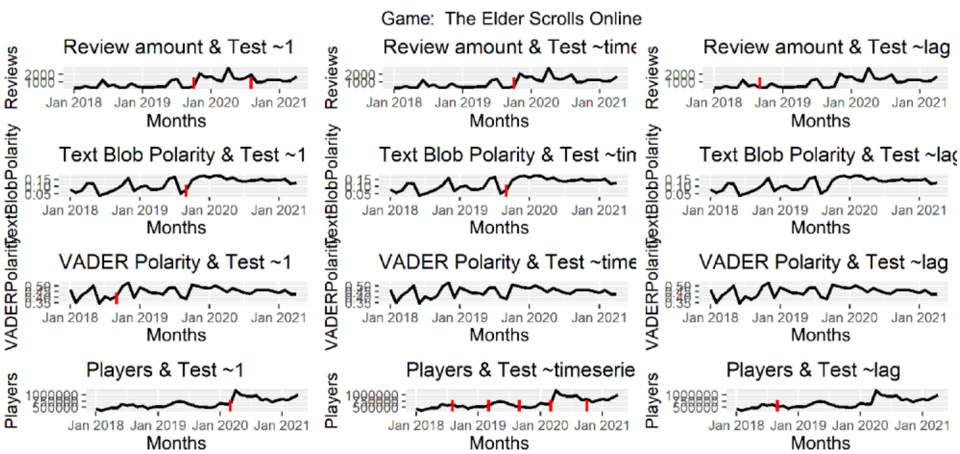

Figure 20 - Structural break test results: The Elder Scrolls Online

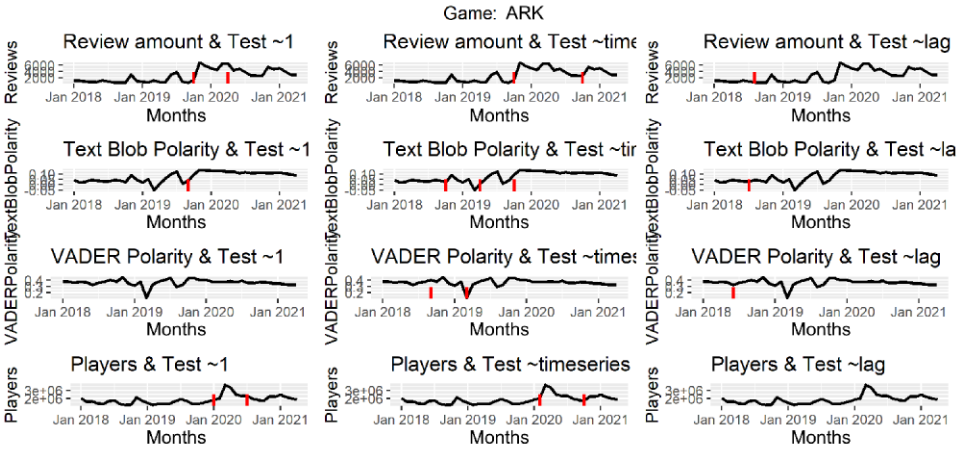

Figure 21 - Structural break test results: ARK

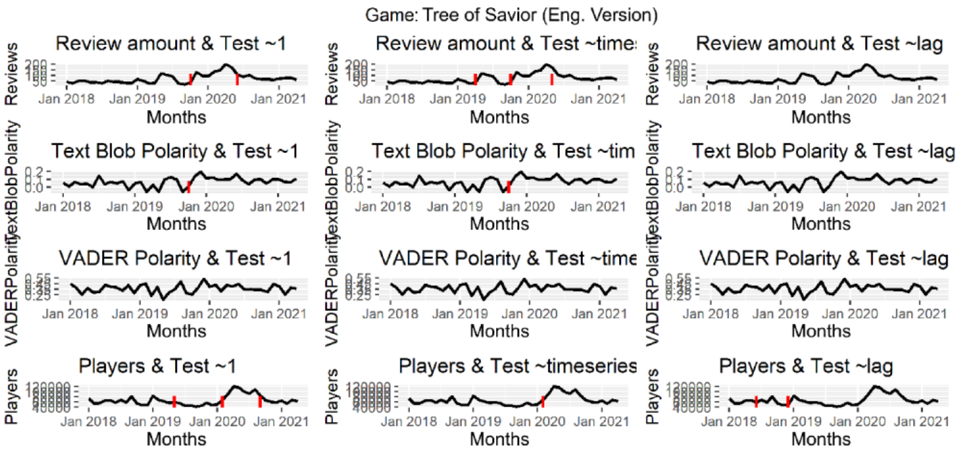

Figure 22 - Structural break test results: Tree of Savior (Eng. Version)

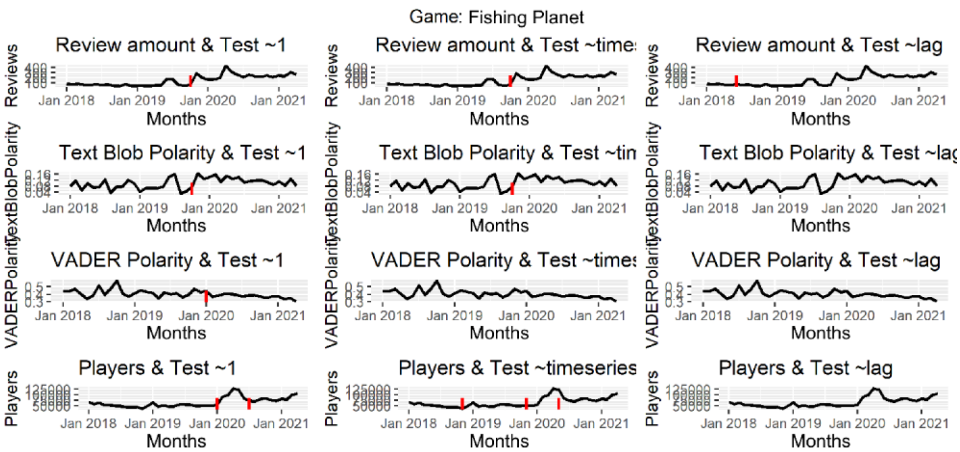

Figure 23 - Structural break test results: Fishing Planet

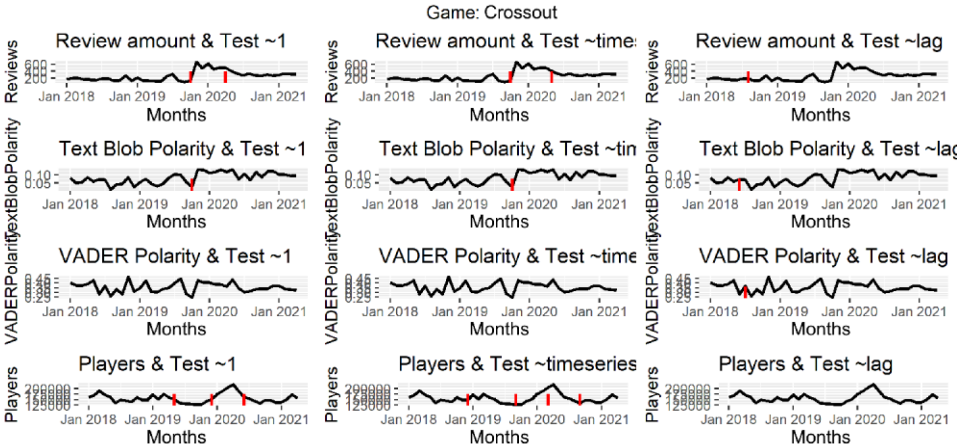

Figure 24 - Structural break test results: Crossout

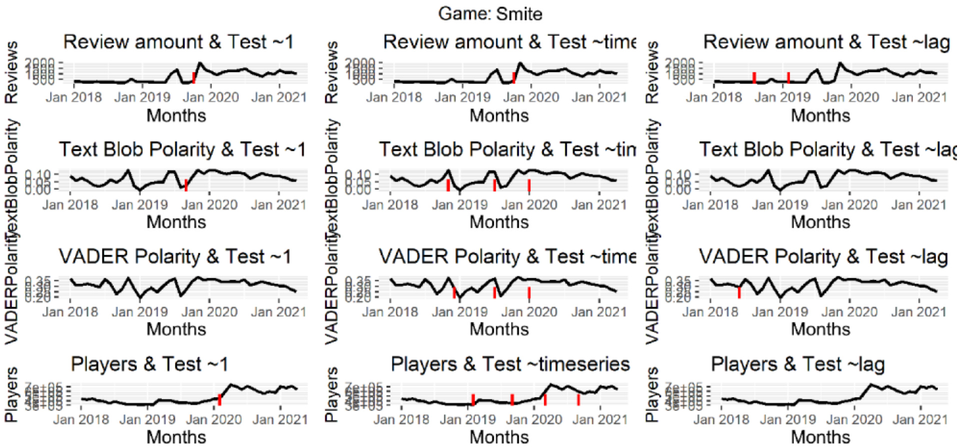

Figure 25 - Structural break test results: Smite

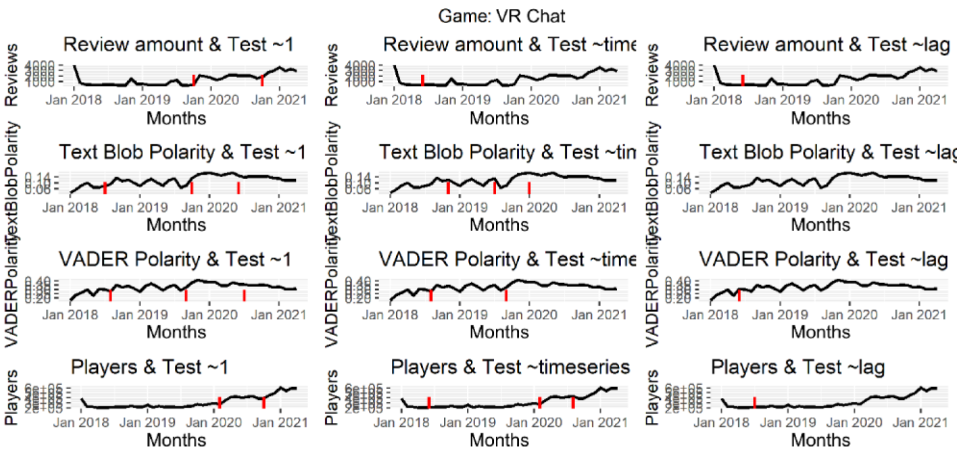

Figure 26 - Structural break test results: VR Chat

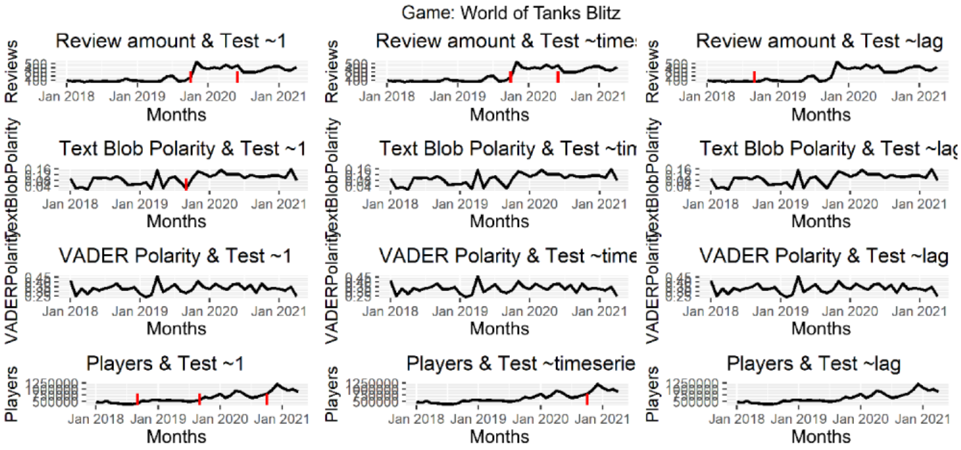

Figure 27 - Structural break test results: World of Tanks Blitz

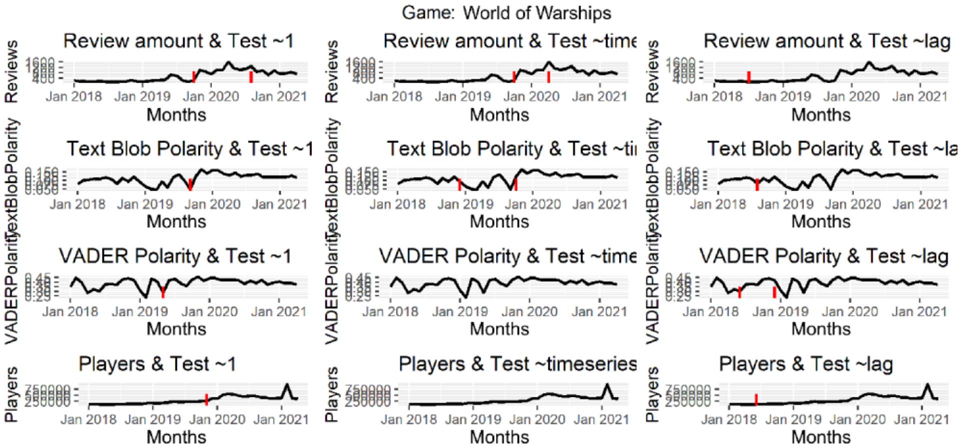

Figure 28 - Structural break test results: World of Warships

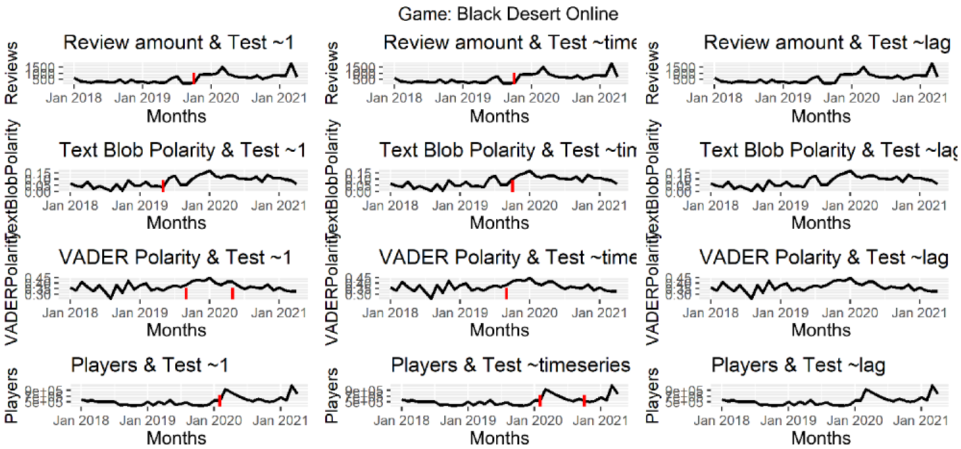

Figure 29 - Structural break test results: Black Desert Online