MBD Project Report - Group 9
Analysis of game reviews and tweets

L. Klunder, G. Georgieva, V. Pirinski, J. Huh

January 2024

Abstract This study explores the interplay between social media activity and user engagement in the context of the gaming industry, by analyzing Steam game reviews and Twitter discourse. Utilizing data from January to October 2020, this work employs statistical methods such as Pearson correlation, Granger causality, and cross-correlation with lags to investigate the (temporal) relationship between the volume of game reviews and the this of game-related tweets. The findings reveal event-driven peaks in user engagement on both platforms, with a notable weak negative linear relationship overall. Specific game analysis, such as the significant updates of Terraria and the PUBG ban in India, demonstrates the impact of external events on digital word-of-mouth. The correlation between playtime and review sentiment on Steam is assessed, offering insights into user satisfaction. Additionally, the study examines the potential presence of a power law distribution in the number of reviews posted by users. The research provides valuable implications for stakeholders in game development, marketing, and platform management. Limitations include the focus on five all-time popular games, which may not represent the entire gaming landscape, and methodological constraints such as keyword filtering and certain data omissions. Future work could extend the temporal analysis, include a larger dataset, and refine data processing methods to enhance the understanding of user behavior in the digital gaming domain.

1 Introduction

In recent years, the gaming industry has witnessed a substantial increase in user engagement, particularly during the COVID-19 pandemic in 2020 [13][16]. This increase in gaming activity has occurred alongside general trends in consumer behaviour. Users are increasingly researching and forming opinions online before investing their time and resources into a game [3] [20]. This phenomenon underlines the importance of platforms like Steam and Twitter in shaping public opinion and influencing gaming trends, as they provide accessible platforms for gathering information.

Steam is a predominant digital distribution platform for video games. It offers unique insights into consumer preferences through its review system. These customer-reviews serve not only as a guide for new potential players but also provide feedback to game developers.

The process of leaving a review on Steam involves users expressing their opinion about a game after fulfilling certain eligibility criteria, such as having purchased the game. They can rate the game and write a detailed review, which is subject to community guidelines and moderation. This feedback mechanism is crucial for understanding player satisfaction and areas for improvement, as reviews can be filled with feedback and critique [1].

The steam platform has two major components namely the Steam Store where you can buy the games and the Steam Community. Steam offers the highest number of PC games compared to other PC gaming distribution platform [11].

Similarly, X, also known as Twitter, with its expansive user base has become a significant platform for discussions about video games. The platform's widespread use make it an ideal place to measure public sentiment and trends about various games [19]. This study aims to delve into the behaviour of users on both Steam and Twitter. It seeks to understand if there is any correlation in the frequency and nature of reviews on Steam and tweets about a particular game. Such an analysis can reveal if one platform's activity influences or reflects the other, offering a more comprehensive understanding of user feedback dynamics.

Moreover, the research aims to compare not just the quantity but also the quality of reactions on both platforms. Understanding whether users express positive or negative sentiments can provide insights into how the game is received.

An essential aspect of this study is the examination of playtime as a metric. Playtime at the time of review and total playtime accumulated, can be a critical indicator of a game's appeal. Understanding the correlation between playtime and review sentiment can offer insights into user satisfaction, potentially influencing future game development. This metric can provide quantifiable data on how much players actually engage with a game.

In this study, we also explore the application of the power law to user review patterns within the gaming industry. The power law, a statistical relationship indicating that a small number of occurrences are extremely common while large occurrences are rare, suggests patterns in user engagement and feedback. In the context of game reviews on platforms such as Steam, this translates to a few users being responsible for a large proportion of the reviews, while the majority contribute sparingly.

Stakeholders, including game developers, marketers, and platform operators, stand to benefit significantly from this research. For game developers, insights into playtime and review sentiment can guide game design and update cycles, ensuring that games are aligned with player preferences and feedback. Marketers can leverage these insights to tailor their strategies, targeting the most engaged user segments. Finally, platform operators like Steam and Twitter can use this information to enhance user experience, recommend games more effectively, and foster a more engaged community.

The research presented here aims to fill a critical gap in understanding the interplay between user behavior on digital gaming platforms and the broader gaming community. By exploring the relationship between various user engagement metrics, like on Steam and Twitter, this study seeks to contribute to a deeper understanding of digital consumer behavior in the gaming industry.

The aim of this study is thus to answer the following questions:

1.1 Research Questions

- Is there a correlation between the daily influx of game reviews on Steam and the concurrent tweets on these games from January to October 2020?
  - What are the visible anomalies and trends over time in Steam reviews and corresponding tweets?
  - How do general time series correlation metrics, such as Granger causality and the Pearson coefficient, reflect the relationship between Steam reviews and tweets?
  - Is there a closer correlation observed when comparing only positive and only negative Steam reviews and tweets?
- What is the correlation between playtime and the recommendation status of Steam reviews?
  - Does the amount of playtime at the time of writing the review have a significant influence on whether an individual recommends the game in the review or not?
  - Is there an observable impact of the recommendation status of Steam reviews on the total playtime after the review? Specifically, do individuals who do not recommend the game continue to play it after expressing their opinions in the review?
- Is there a power law distribution in the number of reviews posted by Steam users?

2 Related work

The relationship between social media activity and product sales has been examined by multiple studies, with the majority of them consistently confirming a correlation between the impact of user-generated content, and specifically social media word-of-mouth (WOM), and consumer behaviour.
One such example is the work of Rui et al.[15] that provides foundational insights into how Twitter activity influences movie sales by demonstrating that the volume and sentiment of tweets can significantly impact consumer purchasing decisions, with positive WOM correlating with increased sales and negative WOM with decreased sales. This relationship emphasizes the potential of social media, specifically the role of Twitter, as a predictive tool for product success in the market.
Further expanding on the digital landscape's dynamics, Heiens & Narayanaswamy [8] explore how social media endorsements, herein Facebook "likes" and Twitter tweets, contribute to web sales and website visits. They conclude that social media is capable of enhancing product visibility and appeal through digital social legitimization. Jones et al.[5] further add to this body of knowledge by focusing on the "Twitter effect" on service performance. Their study indicates that Twitter serves as an essential platform for disseminating post-purchase evaluations, thereby influencing early product adoption and sales. Tang [18] further explores the predictive potential of Twitter, demonstrating its capacity to not only reflect but also influence sales growth. Evidently, research positions Twitter comments as an active component in shaping sales dynamics, extending beyond mere reflections of consumer sentiment.
This work aims to thus extend these findings into the realm of video games, particularly examining the correlation between game reviews on Steam and concurrent game-specific tweets, and whether trends and/or abnormalities in one, e.g., in the form of a spike or a plateau, have a spill-over effect on the other. To the team's knowledge, no such research in this particular field has been conducted at the time of writing this paper.
To better depict the role of game reviews online, the team bases its platform choice on several building papers as follows. This of Phillips et al. [14] provides a significant contribution to the understanding of user feedback within the gaming community. Their research underscores the depth of insights that can be gleaned from user interactions on gaming platforms, highlighting the influence of these platforms on game perception and game sales success. Furthermore, Cox & Kaimann [4] provide a comparative analysis of the influences exerted by professional critics and user-generated reviews on video game sales. Their study reveals a complex dynamic wherein user-generated content can sometimes overshadow the impact of professional reviews, signifying a shift in the traditional paradigms of marketing and consumer influence.
Specifically with regards to Steam, previous work shows that with an increase in playtime, the player is

more likely to recommend the game [6], with a further paper pointing out an average of 6.76 hours of game playtime for a positive review to be written in comparison to negative ones that on average take 2.06 hours of playtime to emerge [7]. This work aims to explore this further, as to dismantle or support these findings, by increasing the dataset examined, and by selecting varying game categories to tackle bias.
Despite the addressed advancements, there remains a research gap in understanding the correlative dynamics between different types of user interactions across social media platforms and gaming review sites, particularly in a comparative temporal context. This study aims to address this gap by methodologically exploring the relationship between game reviews on Steam and corresponding Twitter activity within a specified timeframe. Moreover, the investigation into the correlation between in-game playtime and the sentiment of reviews on Steam, as well as the exploration of potential power law distribution in user review behavior, aims to contribute deeper insights into user engagement patterns and their implications for game development and industry practices. The subsequent section will detail the data sources used to do so.

3 Data

3.1 Twitter

In this project, the Twitter dataset spanning from January to October 2020 is used. This dataset, stored in the Hadoop Distributed File System (HDFS) with the path '/user/s2465795/project_data/twitter/' directory, comprises approximately 65GB per month, with notable deviations in February and March, recording a high volume of 415GB and 104GB respectively. The extensive data volume in February led to variations in the folder structure, resulting in the exclusion of numerous files from consideration in the project.

For the project, the data was filtered such that only non-empty tweets would be considered and two columns, 'created_at' and 'text', were selected and used. The column 'created_at' was initially in the format datetime, which was converted to date format (month/day/year) for the daily temporal analysis. This column shows timing information to aid the understanding of when tweets were written. The column 'text' serves to enable the sentimental analysis and identification of discussions related to various games.

To properly identify the tweets specifically discussing the games of interest, a keyword-based approach is employed. Game names and their abbreviations serve as keywords, enabling the filtering and extraction of relevant tweets discussing the games from the dataset. This led to around 15MB of data with columns date, game_name, and recommendation status on whether positive or negative review. This was processed locally for the visualization of graphs.

3.2 Steam game reviews

The Steam game reviews dataset from 2006 to January 2021 encompasses around 8GB. The dataset, located in HDFS at the path '/user/s2465795/project_data/', contains review text with the details on associated information of the authors' Steam accounts. For the project, only data from January to October 2020 is used to match with the tweets data, which is approximately 800MB.

The columns selected to analyze Steam game reviews are 'app_name', 'timestamp_created', 'recommended', 'author.num_reviews', 'author.playtime_at_review', and 'author.playtime_forever'. The 'app_name' column identifies the game name of the corresponding review context. The column 'timestamp_created' was converted from UNIX timestamp to date format (month/day/year) to increase the readability of the data and the temporal analysis together with the tweets data. The 'recommended' column denotes whether the review recommends the game or not, contributing to the exploration of correlation with playtime. Additionally, author-related information, such as 'author.num_reviews' and 'author.playtime_at_review', allows for a deeper understanding of the reviewer's profile, contributing to the potential insights into user behavior. The 'author.playtime_forever' column, in particular, sheds light on the cumulative playtime of the reviewer, providing valuable context for the analysis of the correlation between playtime and recommendation status.

3.3 Comparative data

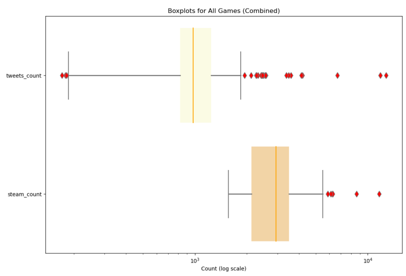


Figure 1: Daily Boxplots of Steam and tweets count for all games in log scale

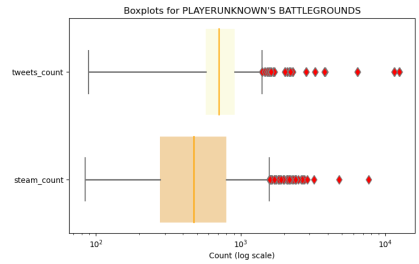


Figure 2: Daily Boxplots of Steam and tweets count for PUBG in log scale

The boxplot presented in Figure 1 illustrates the total daily counts of data, post-filtering the data. Upon examining the boxplot across all games, it becomes evident that the median total count of Steam reviews notably exceeds that of tweets daily. This is consistently observed when analyzing the counts per game per day across the majority of games.

Specifically, for tweets, the median is lower than that of Steam, featuring a shorter left box of the median line compared to the right box. The left whisker line for tweets is longer, and the presence of numerous outliers on the right side is evident, with a few outliers on the left.

In contrast, Steam has a higher median value than tweets, with the left box of the median line being larger than the right box. The left whisker line is relatively shorter with a marginal difference for the right whisker line. Outliers for Steam are concentrated on the right side of the boxplot, with a lower occurrence of outliers compared to tweets.

In contrast to the general trend observed across the selected games, Figure 2 for PlayerUnknown's Battlegrounds (PUBG) shows a unique pattern. The median count of Steam reviews for PUBG appears lower than the median count of tweets per day.

The boxplot for tweets shows the symmetric width of the boxes on both sides of the median line. The whisker lines for tweets are longer on the left, and outliers are concentrated on the right side. Steam reviews have wider boxes compared to tweets while the boxes also show symmetry on both sides of the median line. Although the left whisker line is longer than the right, the difference is less pronounced than that observed for tweets. Steam also has outliers on the right side of the boxplot.

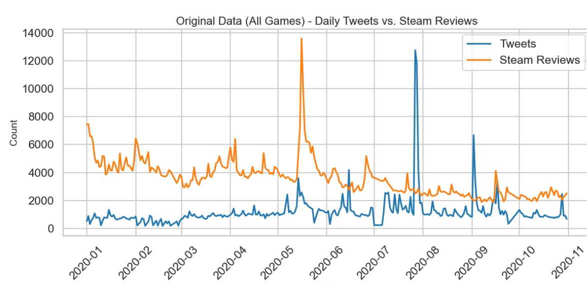


Figure 3: Line graph on Daily counts of Steam and tweets for all games

The line graph in Figure 3 shows the daily counts of both Steam reviews and tweets for all games on how much user engagement has been over the observed period. Notable peaks in activity are evident, particularly a significant surge in Steam observed on May 16, 2020, and pronounced peaks in tweets in July and early September 2020. For the May 16, 2020 peak, it is not only Steam that has the peak but also tweets. It shows a sudden increase and subsequent drop in the same period. This is due to the major update in the game Terraria [17], reflecting a shared surge in user interest across both platforms. Another peak in tweet activity observed in July corresponds to the banning of the game PUBG in India after an incident involving a teenager using the life savings of his father [9] . Despite these distinctive peaks, the overall trend indicates that tweets consistently exhibit lower daily counts compared to Steam reviews, with only occasional exceptions. The correlation between Steam and tweet activity during specific events shows the interconnected nature of user engagement in the data, where noteworthy gaming developments or controversies can lead to increased discussions on either both or one of the platforms.

4 Methodology

4.1 Steam reviews

4.1.1 Playtime to recommended correlation

As explained previously, playtime can be a valuable indicator when evaluating people's opinion of games. To explore its possible correlation with whether a review is a recommendation or not, this study will make use of a boxplot of the distribution of the playtime for each user posting a review, at the time of the review, as well as the total playtime at the time the dataset was constructed. This is by no means a statistically significant method, but it aims to visualize the data to gain insight into whether it could be worth exploring in the future.

Furthermore, doing this for the playtime at both the time of the review and the total time can provide some data about the behaviour of players until and after posting a review on Steam. For example, one would expect that people who do not recommend a game to have quit earlier, and therefore have lower playtime at the time of writing the review, and, naturally, that those who recommend the game would play it longer on average after the review than those that do not recommend it.

4.1.2 Power law in the number of reviews posted by users

Another potentially useful metric that was mentioned earlier is the number of reviews a user has posted. This can be used to partially understand the reviewing behaviour of users. Those who have posted many times might have a different pattern of behaviour than those who post infrequently. But if a substantial part of the reviews comes from users with a given frequency of reviews, this would mean that the data is heavily skewed and the opinions of the smaller groups may get lost in the analysis of the general opinions.

This paper will explore the potential presence of power law in the distribution of the number of reviews posted by individual users. Power law is a statistical concept describing the relationship between two variables, indicating that majority of the reviews come from a small group of users. To check for this property, a normal scale plot of the number of users with a given number of reviews against the total number of reviews they have, accompanied by a log-log plot of the same data. Observing a straight line on the latter plot is the standard method of checking for the presence of power law.

4.2 Comparison of Steam reviews and Tweets

Two key statistical techniques - normalization and moving averages - are initially employed to prepare and analyze the data. Said techniques are crucial for making meaningful comparisons and drawing reliable conclusions from the data.

Normalization. After plotting the initial raw counts of data, it was established that without normalization, direct comparisons of this data could be misleading due to the differences in the magnitude of raw counts, as displayed in Figure 3. This could be due to inherently different user bases across the two platforms, and/or the platform-specific engagement patterns.

A MinMaxScaler is utilized that transforms each feature (count of tweets and reviews) to a scale between 0 and 1, making them directly comparable.

Moving Average. In this work, Moving averages are used to smooth out the short-term fluctuations in data, and to highlight longer-term trends in the time-series data. A 7-day moving average is calculated for both the normalized tweet counts and review counts. This helps in reducing the impact of transient spikes or drops in daily counts, which might be due to temporary events or anomalies rather than representing a sustained trend. By averaging data points over a week, some noise filtering is supported without smoothing out too much data.

4.2.1 Pearson coefficient

The Pearson Coefficient measures the linear correlation between two variables, herein the two sets of data. In this study, it is used to assess the strength and direction of the relationship between the normalized daily counts of Steam reviews and the daily number of tweets on Twitter, both overall and game-specific. A high positive Pearson coefficient here would suggest that as the activity on one platform increases, the activity on the other tends to increase as well, indicating a potential interactive relationship between Steam reviews and Twitter discussions.

4.2.2 Granger causality

The Granger Causality test is employed to examine whether changes in one time series can predict changes in another. In this context, it helps determine if the number of Steam reviews can predict the volume of Twitter activity concerning the same games, or vice versa. This method is valuable for identifying potential lead-lag relationships between the two platforms, suggesting which platform might be driving the conversation about a game.
In this test, lags represent the number of time periods that are looked back on in one time series to predict another. In this case, a maximum lag of 30 days is chosen to encompass a broad enough timeframe to capture potential delayed effects in the data, which is important for understanding the dynamics of online discussions and reviews that may not be immediate. As the examined data spans over 10 months, a lag of 30 days enables the team to capture and analyze the effects that might occur within a monthly cycle without concerns about the scarcity of data, which could be an issue in shorter timeframes.
The p-value of a Granger causality test represents the probability of finding the causality that was found if there was no causality present in the data. A $p - value < 0.05$ indicates that there is likely some Granger Causality between the two time series for the given load. If the p-value is less than 0.05, it means that the values of one time series predict the values of another time series in the future. This is used as the exclusion criteria for representing Granger Causality graphs across the examined games in this work, meaning games with a $p - vaue > 0.05$ are not be considered relevant for this criterium, and thus not discussed in the Results part, unless they represent extreme outliers.

4.2.3 Cross-correlation plus lag

The conducted Cross-Correlation plus Lag analysis explores the relationship between two time series at different time intervals (lags). This method helps to identify if there is a delayed (lagged) effect of the activity on one platform on the other. For instance, an increase in Steam reviews may correlate with an increase in Twitter activity after a certain number of days. In this context, it is applied to assess how Steam reviews (as reflected in their normalized count) and Twitter activity regarding specific games are related across different time lags. The function computes the cross-correlation coefficients for lags ranging from -30 to +30 days. Understanding these lag effects can provide insights into how user feedback propagates across different platforms and over time. The choice of maximum lag being 30 is consistent with the Granger Causality analysis.
In this study, the main difference between the two statistics is that the Granger Causality can identify if trends in Steam reviews are predictive of trends in Twitter discussions about games, or vice versa, within a 30-day lag whereas the Cross-Correlation with Lag analysis will show how these two series are correlated (move together) at different time intervals without implying that one causes or predicts the other.
Together, these methods provide a comprehensive understanding of the temporal dynamics between Steam reviews and Twitter activity, combining predictive insights from Granger Causality with a broader correlation perspective from Cross-Correlation.

4.2.4 Boxplots for recommended vs non-recommended reviews/tweets

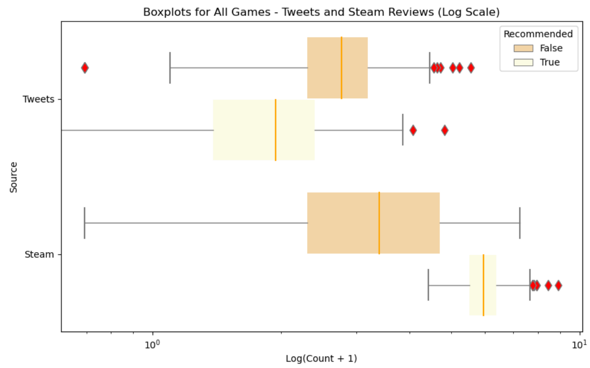


Figure 4: Daily Boxplots of recommendation status on Steam reviews and tweets count for all games in log scale

The boxplot in figure 4 compares the daily recommendation status (positive or negative) for both Steam reviews and tweets across all games. This provides valuable insights into the distribution and characteristics of user feedback. A negative recommendation means a negative Steam review or negative tweets while a positive recommendation shows a positive Steam review or positive tweet posting about a specific game.

For tweets, false recommendations exhibit a higher median than true recommendations, and the false box is shorter than the true box. The symmetry between the lower edge of the false box and the higher edge of the true box suggests similar interquartile ranges. However, the left side of the true box is notably longer than the right, indicating a skewed distribution towards lower values. This suggests that for true recommendations in tweets, there is a tendency for a relatively higher number of days with a lower volume of reviews. This does not necessarily mean there are fewer positive reviews overall, the count of positive reviews is more concentrated at lower levels on the log scale on many days.

On the other hand, the false recommendation for Steam reviews has a median higher than both true and false medians in tweets. This also has the longest box among all the other boxplots, reflecting a wider spread of data. The symmetry in box sizes for false recommendations in Steam suggests a balanced distribution. Additionally, both true and false recommendations in tweets and false recommendations in Steam have longer whisker lines on the left, emphasizing an extended spread of lower values. In contrast, the true recommendations of Steam reviews feature the highest median but the smallest box size among all boxplots. This indicates a more concentrated distribution around the median. There is also a symmetry of left and right boxes for the true recommendation of Steam reviews which further highlights a balanced spread of data.

4.2.5 Sentiment analysis

In this study the tweet text is processed to find the sentiment of tweets about particular games. The tool used is the NLTK VADER sentiment analyzer, which uses a lexicon specifically designed to analyze context in Tweets [10]. There are other python libraries that could be used for this task, such as the open-sourced TextBlob, which can generally be applied for any type of content [12]. Another study has compared their performance on tweets and found that both perform decently, with accuracies of 79% and 73% respectively. However, the former one is used mainly because of easy access to existing documentation on how to import the library for distributed processing of big data, without consideration for their efficacy on tweets about games.

To make the same division of positives and negatives among Steam reviews, no sentiment analysis tool is used. Instead, the attribute, marking if the review recommends the game or not, is used directly. This categorization is determined by the Steam users directly when making the review. Some studies have suggested that there might be discrepancies between user ratings and text reviews [2], so it might be beneficial to check if this is the case for the Steam data. However, this is not within the scope of this study and the recommendation attribute is taken as the absolute truth.

5 Results

The upcoming section examines how Steam reviews and Tweets about popular games correlate and potentially influence each other. The analysis draws on a variety of statistical tools and methods, like the Granger causality and cross-correlation with lags. The results aim to provide clarity on the reliability and significance of the observed trends and correlations.

5.1 Steam vs. Twitter

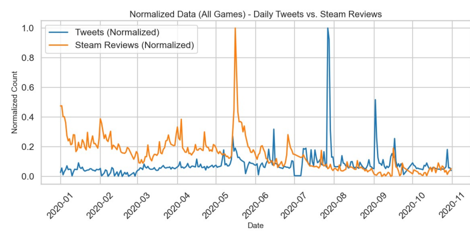


Figure 5: Normalized line graph on Daily counts of Steam reviews and tweets for all games

The normalized line graph as seen in Figure 5 is to compare Steam reviews and Twitter tweets by normalizing all values to fall within a range of 0 and 1. Both lines have multiple peaks, indicating a relatively high number of tweets and reviews. The Steam reviews peak is an event-driven peak as Terraria released a new update resulting in a peak in new reviews as it (re)drew players to the game. The prominent blue spike at the end of July is caused by a PUBG ban in India, causing more Twitter traffic and WOM buzz. The Pearson coefficient is -0.093, indicating a weak negative linear relationship.

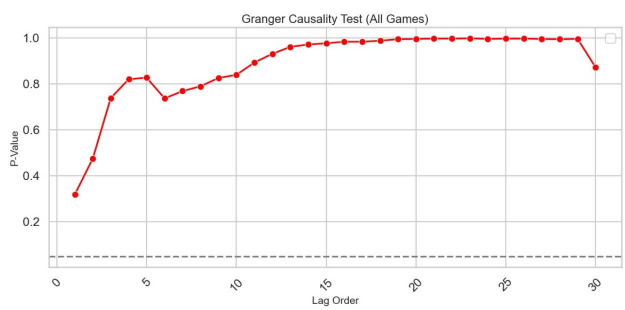


Figure 6: Granger Causality test for Steam reviews and tweets for all games

In Figure 6, the Granger Causality test displays p-values above the 0.05 significance threshold for all lag orders. This indicates that there is no statistically significant evidence to suggest a predictive relationship between the series within the observed lags.

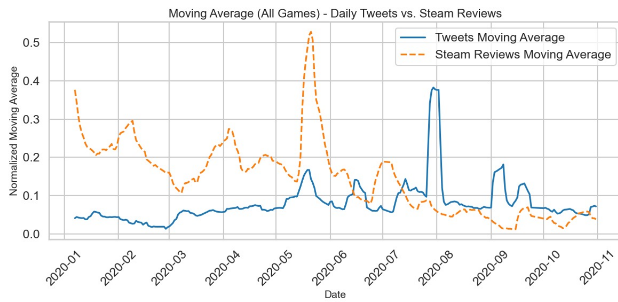


Figure 7: Moving Average for Steam reviews and tweets for all games

In Figure 7 the moving average can be observed. There are two peaks, one in May and one at the end of July. The spikes are likely because of the event with PUBG, in August and the event and the release of a new Terraria update.

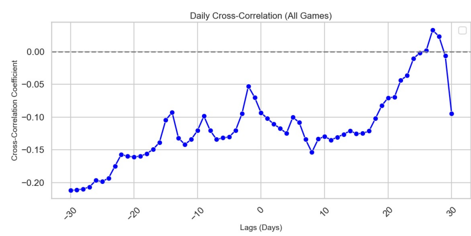


Figure 8: Cross-Correlation daily

In Figure 8 mostly a negative, relationship between the time series is to be observed. This indicates an inverse relationship. The largest peak is at 30 days, indicating that there is a positive relationship between Steam reviews and Tweet data at the thirty-day mark.

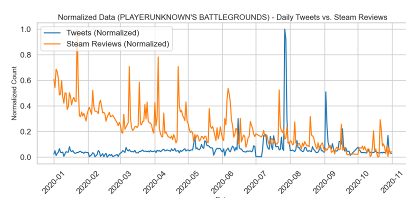


Figure 9: Normalized data for PUBG

PUBG. The correlation coefficient of PUBG daily data is negative with the value -0.1634, suggesting a weak negative linear relationship between the daily counts of Steam reviews and tweets. Additionally, the magnitude of the correlation is relatively small showing the weak strength of the linear relationship.

There are some fluctuations visible in Figure 9, which matches with the correlation coefficient suggesting the slight tendency for one variable (Steam reviews or tweets) to decrease as the other increases or vice versa, but with no strong relationship.

Among those few fluctuations, there is one very notable peak in tweet data in July 2020. This peak corresponds to a specific incident of a teenager using the life savings of his father to make in-game purchases in PUBG. The significant attention gained on this incident led to a surge in discussions and reactions on social media platforms, particularly on Twitter. The sudden increase in tweets compared to Steam reviews during that period can be attributed to the widespread sharing of news, opinions, and discussions on this incident. This likely caused the spark in PUBG-related tweets causing temporary deviation from the usual trend of more daily Steam reviews than tweets. The sensational nature of the news, along with the public's interest and engagement with the story contributed to the spike in tweets during that specific time frame, showing how external events and controversies can impact the online discourse around a game, leading to notable fluctuations in data patterns.

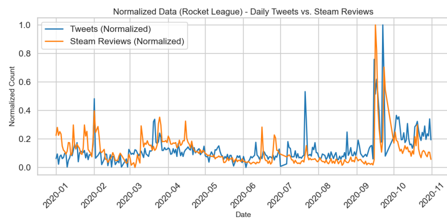


Figure 10: Normalized data for Rocket League

Rocket League. The analysis of Rocket League daily normalized data on Steam reviews and tweets shows a moderate positive linear relationship between the variables, as indicated by the Pearson correlation coefficient of 0.4893. This coefficient suggests on average, as one variable increases, the other tends to increase as well, and vice versa. This trend is demonstrated in Figure 10 with substantial overlap, indicating the high degree of similarity or synchronization in the patterns of the Steam reviews and tweets over time.

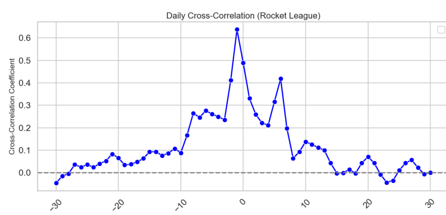


Figure 11: Daily cross-correlation for Rocket League

The cross-correlation graph for Rocket League in Figure 11 with lags ranging from -30 to 30 shows that the cross-correlation coefficients start a little below 0.0 and generally increase as they approach the central lag of 0. This indicates a tendency for one variable to follow the other with a time delay.

The fluctuation in the cross-correlation coefficients suggests that the relationship between the variables is not static and may be influenced by various factors or time-dependent dynamics in the Rocket League data. The periods where the two variables exhibit a stronger positive relationship (coefficient increases) indicate synchronization in their behavior with a time lag. Conversely, the periods where the coefficients decrease may signify a temporary divergence or a shift in the synchronization pattern.

The absence of cross-correlation coefficients significantly lower than 0 suggests that there is no clear evidence of a consistent negative relationship between the variables at different time lags.

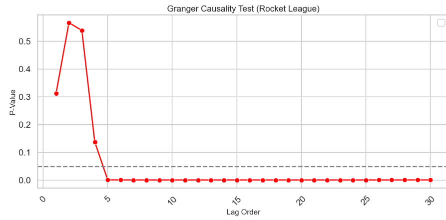


Figure 12: Granger Causality for Rocket League

The Granger causality results for Rocket League in Figure 12, as plotted with 30 lags, indicate interesting temporal patterns in the relationship between the variables. The presence of plots higher than 0 for lags 0 to 4 suggests evidence of Granger causality during these early lag periods. This implies that past values of one variable provide information about predicting the other during these specific time lags.

However, the intriguing aspect is the abrupt drop to 0.0 for lags 5 to 30, implying a lack of statistically significant Granger causality beyond the initial few lag periods. This sudden drop suggests that past values of one variable cease to provide useful information for predicting the other after the initial lags. This temporal pattern provides insights into the dynamic relationship between the variables and suggests that the influence of past values on prediction is more prominent in the short term than in the longer term.

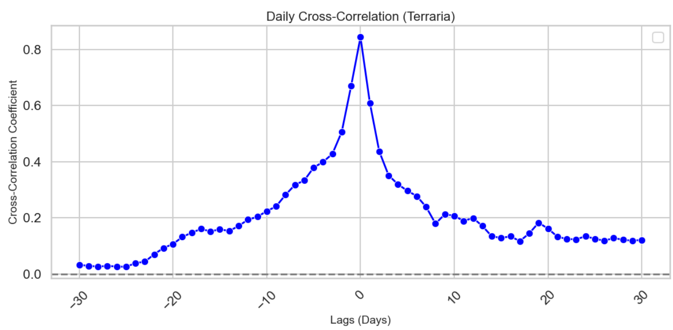


Figure 13: Cross-Correlation (Terraria)

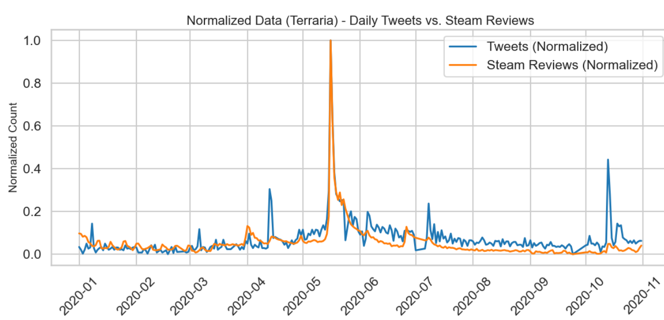


Figure 14: Normalized Data (Terraria)

Terraria. The Pearson correlation coefficient is very high at 0.844 (the highest examined out of the five games subject to this study), indicating a strong positive relationship between Twitter mentions and Steam reviews. This suggests that changes in one of these variables are closely associated with changes in the other.

With regards to the normalized plot, there is one significant peak in May that aligns with the release of Terraria's final major update, Journey's End, launched on May 16, 2020. This significant update is said to have been highly anticipated and has brought a substantial amount of new content to the game, naturally sparking discussions on Twitter and resulting in many new reviews on Steam, all contributing to the online WOM peak.

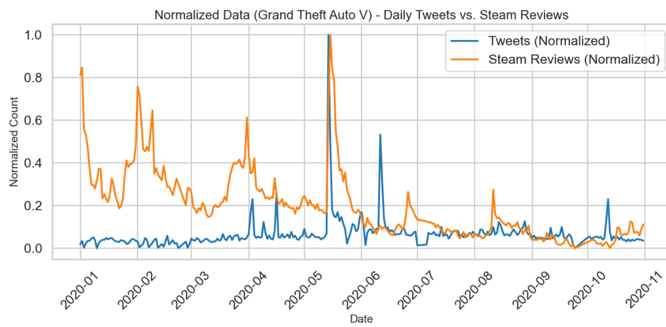


Figure 15: Normalized Data (Grand Theft Auto V)

Grand Theft Auto V. The Pearson correlation is 0.163, indicating a slightly stronger positive yet still insignificant linear relationship between daily tweets and Steam reviews of the game. When looking at the normalized data, seasonality and several peaks in the data can be observed, resulting from multiple significant events throughout 2020 that explain these variations. Specifically, in January 2020, "Grand Theft Auto V" was added to Xbox Game Pass, which contributed to the peak in activity observed as more players gained access to the game. Furthermore, the "Los Santos Summer Special" update was released on August 11, 2020, which correlates with the spikes observed in the data around that time on both platforms, with Steam leading, as players return to the game to explore new features.

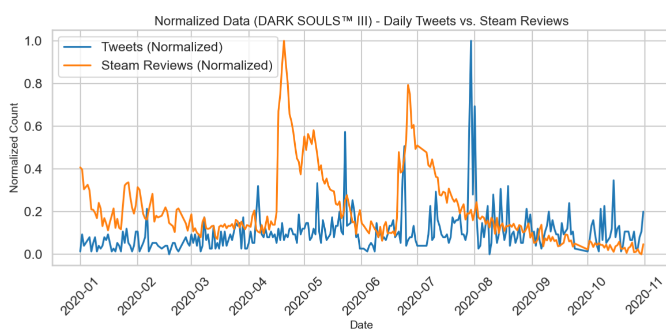


Figure 16: Normalized Data (DARK SOULS™ III)

DARK SOULS™ III. The Pearson correlation coefficient of 0.046 suggests a very weak positive linear relationship between the daily counts of tweets and Steam reviews for the game "DARK SOULS™ III" over the 10 months in 2020. The low Pearson correlation coefficient is also visually confirmed by the lack of consistent simultaneous peaks in the two data sets.

Looking at the normalized data graph, there are visible spikes in the data, which could indicate specific events or releases, e.g., patch releases, related to the game that generated activity on both platforms, yet the team did not manage to pinpoint the exact reasons for this. The Twitter activity appears to have more volatility and higher peaks compared to the Steam reviews, which could indicate a more immediate and short-lived reaction on social media compared to the longer-term trend of reviews being posted on Steam.

SENTIMENT COMPARISON

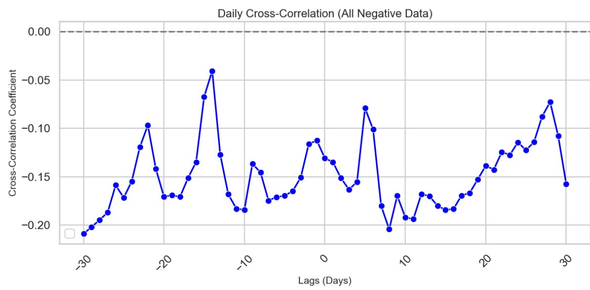


Figure 17: Correlation for negative recommendation

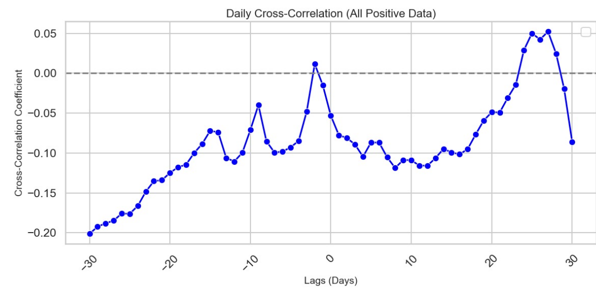


Figure 18: Correlation for positive recommendation

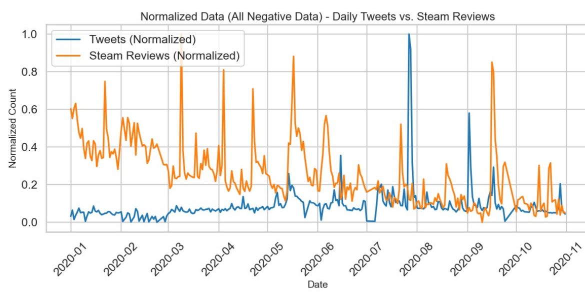


Figure 19: Normalized for negative recommendation

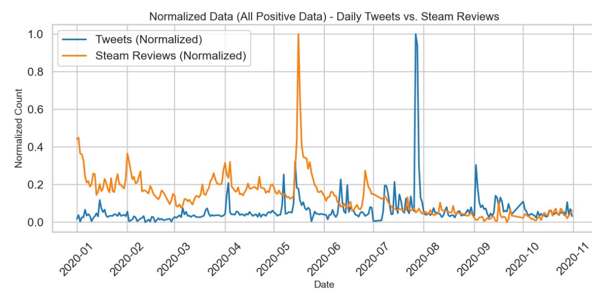


Figure 20: Normalized for positive recommendation

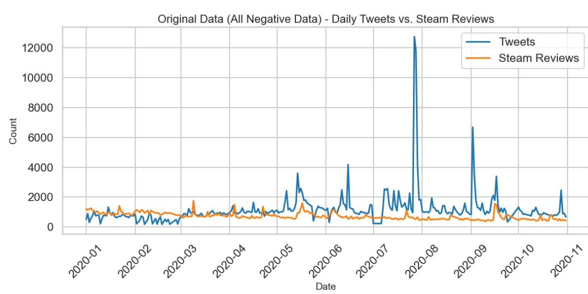


Figure 21: Original data on negative recommendations of all data

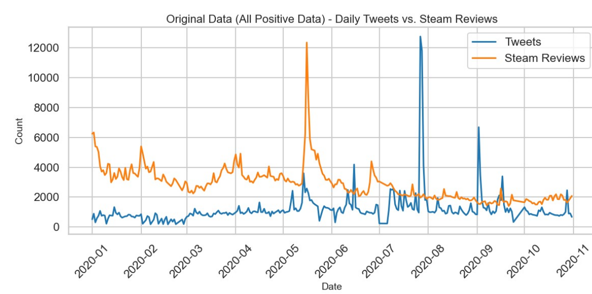


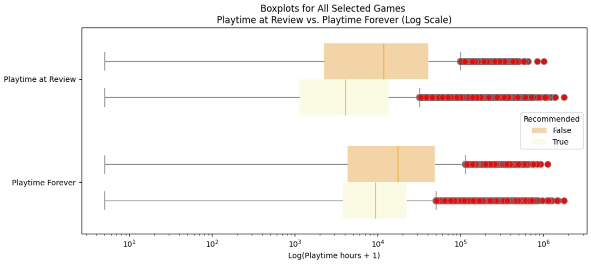
Figure 22: Original data on positive recommendations of all data

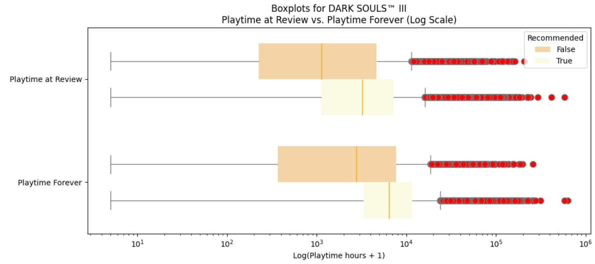
5.2 Steam

5.2.1 Playtime

The plots for playtime are given in the log scale for better visibility. Looking at the overall distribution of the playtimes across the five games, it is evident that the users who do not recommend the games have played them for longer at the time of the review, as well as the total playtime. Interestingly, the median total playtime of users who recommend the games, is lower than the time at review of the users who do not recommend them. This is quite contra-intuitive, but perhaps it is the case that those, who do not enjoy a certain game, end up playing more until they affirm their negative opinion.

Looking at the distributions for the individual games, those for PUBG and Rocket League are similar to the general one. Meanwhile, for GTA 5 there is not much difference between those who recommend and not. However, the graphs for Dark Souls 3 and Terraria are quite different from the rest, but also very similar

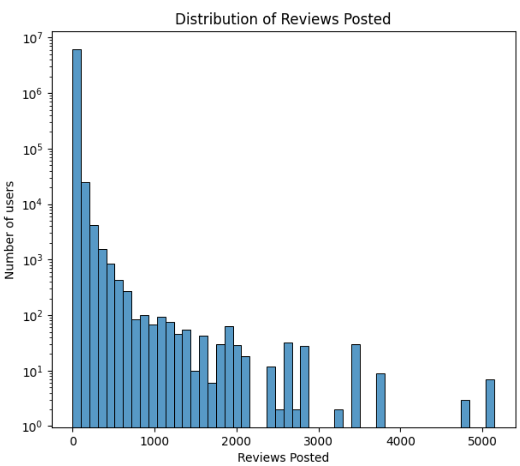


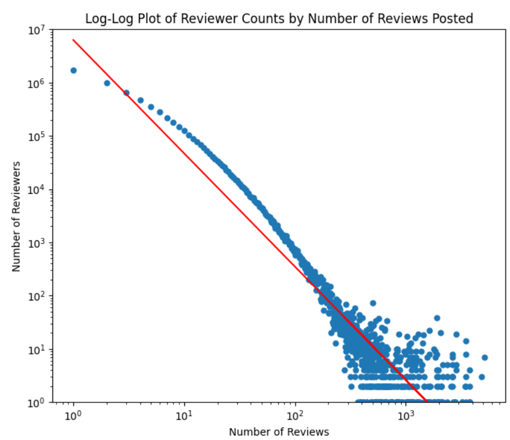


to each other, so only one is shown. For these games the players who recommend the games have played noticeably more and continue to do so after leaving the review. To the extent of the writers' knowledge, the communities of these two games are very dedicated and consistent, whereas a game like PUBG just had waves of popularity.

It is possible that these statistics could be used to describe the behaviour of the player base, and maybe even gain insight into future reviews based on current playtime.

5.2.2 Number of reviews of users





Plotting the number of users with a given number of reviews was unreadable because of the substantial part of people with reviews in the range of a few hundred. Therefore, the y-axis was put to the logarithmic scale, which makes the plot readable and the transformed data looks like an exponential distribution. A log-log reaffirms the notion that the original data does not look to follow the power law, but it is clear that the data is very skewed.

6 Limitation and Future work

6.1 Limitations

The project has certain limitations to consider. First of all, the selection of the five games might not be representative of the gaming landscape, especially for newer or less renowned titles. The chosen games are those already popular and 5 is a small number compared to the available games in the gaming industry. This could affect the generalizability of the findings and limit the applicability of the approach to a broader gaming context.

Moreover, the keyword-based search approach, using game names and relevant abbreviations as keywords, was used to search relevant tweets. Despite the effort made to account for common abbreviations, this approach might fall short, especially for games with names that are common phrases in everyday languages, such as "Among Us". Instead of risking the possibility of including unrelated tweets, the project excluded any keywords that are used in daily conversation. This could lead to the exclusion of valuable data and hinder the comprehensiveness of our analysis.

Additionally, during the processing of February data, the inconsistent folder structure was encountered as mentioned above in section Data. Some data was indented one folder deeper, which was omitted in the results. It was not possible to run the code and extract the data for the full February data since the cluster crashed due to limited memory.

For the Granger causality of Rocket League in Figure 12, the p-value of 0.0 in lags is quite rare and it is consistently 0.0 from lags 5 to 30. This could be due to the precision limitations or rounding in the software. However, due to the limited time, the team was not able to look into the specific cause of this.

6.2 Future Work

Addressing some considerations and exploring additional avenues for improvement can contribute to the refinement and effectiveness of the result and analysis of the project. First, explore hourly data to provide insights into the temporal aspects of gaming discussions, allowing for a more nuanced understanding of trends and correlations. This could be particularly useful in establishing the causality relationships between events and tweets related to gaming.

Expanding the dataset by including the rest of 2020 and January 2021 could enhance the robustness of our analysis, offering a more comprehensive view of the gaming discourse over a more extended period. Generalizing the data processing approach to accommodate different folder structures in Twitter data could address the issues encountered in the data of February, ensuring a more consistent and accurate representation of the collected information. Since the code execution crashed due to the limited memory, a feasible strategy must be incorporated which might involve running the February data in smaller, manageable subsets to prevent data overload and ensure successful execution.

Additionally, examining data around the release period of specific games would help capture the immediate impact and sentiment surrounding the launch of the game, providing valuable insights into the dynamics of gaming communities. Specifically for tweet data, incorporating hashtags to filter the relevant tweets could improve the precision of the data collection or filtering process. This would help overcome the limitations associated with common phrases and abbreviations and ensure a more comprehensive and accurate representation of tweets.

Lastly, it is expected to go over the possible causes of Granger causality consistently being 0.0 in the graph for Rocket League and see whether this came from the no significance of data between Steam reviews and tweets or whether some errors were made, like precision limitations or rounding.

References

[1] Steam reviews. https://store.steampowered.com/reviews/, 2024. Accessed: 2024-01-31.

[2] A. Almansour, R. Alotaibi, and H. Alharbi. Text-rating review discrepancy (trrd): an integrative review and implications for research. *Future Business Journal*, 8(1), Feb. 2022.

[3] T. Chen, P. Samaranayake, X. Cen, M. Qi, and Y.-C. Lan. The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study. *Frontiers in Psychology*, 13, 2022.

[4] J. Cox and D. Kaimann. How do reviews from professional critics interact with other signals of product quality? evidence from the video game industry. *Journal of Consumer Behaviour*, 14(6):366–377, Nov. 2015.

[5] C. Esmark Jones, S. Waites, and J. Stevens. Influence of social media posts on service performance. *Journal of Services Marketing*, 36(2):283–296, 2022.

[6] T. Guzsvinecz. The correlation between positive reviews, playtime, design and game mechanics in souls-like role-playing video games. *Multimedia Tools and Applications*, 82(3):4641–4670, 2023.

[7] T. Guzsvinecz and J. Szűcs. Length and sentiment analysis of reviews about top-level video game genres on the steam platform. *Computers in Human Behavior*, 149:107955, 2023.

[8] R. A. Heiens and R. Narayanaswamy. The impact of digital social legitimization on website visits and web sales. *International Journal of Electronic Commerce Studies*, 12(1):29–42, 2021.

[9] J. Horowitz and S. Gupta. India bans pubg, alipay, and other chinese apps. *CNN*, September 2020. Accessed: 2024-01-31.

[10] C. Hutto and E. Gilbert. 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):216–225, May 2014.

[11] D. Lin, C.-P. Bezemer, Y. Zou, and A. E. Hassan. An empirical study of game reviews on the steam platform. *EMPIRICAL SOFTWARE ENGINEERING*, 24(1):170–207, FEB 2019.

[12] S. Loria. textblob documentation. *Release 0.15*, 2, 2018.

[13] J. Na, E. J. Kim, and J. Kim. Big data analysis of the impact of covid-19 on digital game industrial sustainability in south korea. *PLOS ONE*, 17(12):1–16, 12 2022.

[14] C. Phillips, M. Klarkowski, J. Frommel, C. Gutwin, and R. L. Mandryk. Identifying commercial games with therapeutic potential through a content analysis of steam reviews. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–21, Oct. 2021.

[15] H. Rui, Y. Liu, and A. Whinston. Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, 55(4):863–870, Nov. 2013.

[16] Statista. Number of daily active users of steam worldwide from january 2017 to january 2020. https://www.statista.com/statistics/733277/number-stream-dau-mau/, 2020.

[17] Steam. Terraria: Journey's end launches on may 16, 2020! https://store.steampowered.com/news/app/105600/view/2220779751685536291, April 2020. Accessed: 2024-01-31.

[18] V. W. TANG. Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on twitter. *Journal of Accounting Research*, 56(3):989–1034, Nov. 2017.

[19] Twitter Marketing. How gaming is growing on twitter, 2020. Accessed: [31-01-2024].

[20] A. Yaylc and M. Bayram. E-wom: The effects of online consumer reviews on purchasing decisions. *International Journal of Internet Marketing and Advertising*, 7, 01 2012.