Erasmus University Rotterdam

Erasmus School of Economics,
Department of Business Economics - Marketing Master Thesis

The effect of online User-Generated Content on video game sales

Author:
Vincent Lansbergen

Supervisor:
Mr. F.A. Deutzmann

December 9, 2015

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Abstract

The purpose of this research is to identify if there is an effect of online User-Generated Content (UGC) on the sales of video games. The UGC that collected is specified to games launched for the PlayStation 4 in 2014 with the genre 'action'. We study a period of 8 weeks after the release of the video game. We defined online UGC into four variables: *Volume*, *Negative sentiment*, *Positive sentiment* and *Rating*. We have not been able to retrieve the variables *Negative sentiment* and *Positive sentiment* directly from the data, for which we performed a sentiment analysis.

The sentiment analysis we performed is based on the Naive Bayes method. This model assigns a sentiment to each review, which could be negative, neutral or positive. To analyse our data we use panel data models. We use the Hausman test and the Mundlak test to determine which panel data model is most consistent with our data. Results indicated that the Fixed Effects model was consistent according to both tests. Observations from the results of these models showed that the variables *Volume* and *Positive sentiment* have a positive significant impact on sales. The variables *Negative sentiment* and the *Rating* do not have a significant effect on sales. For our dataset online UGC does have an impact on sales of video games in the first 8 weeks, but further research has to be done before implementing strategies based on these findings.

Contents

1 Introduction **3**

1.1 Problem Statement and Research Objective . . . 4

1.2 Scientific and Managerial Relevance . . . 6

1.3 Structure of the Thesis . . . 7

2 Theory and Hypotheses **7**

2.1 Literature review . . . 7

2.1.1 Defining UGC . . . 7

2.1.2 Methods for classification . . . 10

2.2 Research approach . . . 14

2.3 Dependent Variable . . . 14

2.4 Key Independent Variable(s) and their Relation with the Dependent Variable(s) . . . 14

3 Data **16**

3.1 Empirical settings: the video game industry . . . 17

3.2 Online UGC data . . . 19

3.3 Sales data . . . 20

3.4 Data Collection . . . 22

3.5 Definition of Measures . . . 22

3.6 Data Description . . . 23

3.6.1 Sales data . . . 23

3.6.2 UGC data . . . 27

4 Research Methodology **30**

4.1 Sentiment Analysis . . . 30

4.2 Panel data models . . . 33

5 Analysis and Results **36**

5.1 Sentiment Analysis . . . 36

5.2 Correlation and Munticollinearity . . . 42

5.3 Panel data model . . . 43

5.3.1 Fixed or Random Effects model 45
5.3.2 Fixed Effects model 46

6 Conclusion **51**
6.1 General conclusion 51
6.2 Academic Contribution and Managerial Implications 52
6.3 Limitations and Directions for Future Research 54

A Released Action games 2014 **55**

B Data Extraction **56**

C Process in RapidMiner **57**

D Process sentiment analysis **57**

E Random effects model **60**

1 Introduction

User-Generated Content (UGC) has and will become more important for consumers in deciding whether or not to buy a video game. According to eMarketer 2009, already for 2010, estimates were that 131.4 million users would generate 95.3 million UGC. Reasons for the fast growth of UGC are generally the low threshold, wide reach, low (or no) cost, availability and accessibility. Consumers inform each other on their experience of the product, which can be created at every moment and, thus, is continuous data. For companies, it is almost necessary to act on these evolvements. Strategic decisions must be made to improve sales and customer satisfaction. According to Brandweek 2003, 67% of consumer purchased goods is based on UGC. For this reason, our goal is to study if sales of video games are influenced by this online User-Generated Content (UGC). There already are some papers on this topic for the movie industry, but for video games, literature is limited. Our research question will, thus, be:

RQ: How does online UGC affect the sales of video games in the first 8 weeks after its release?

We define online UGC as volume of UGC and sentiment. First, we will determine the sentiment, by using online reviews on Amazon.com and performing a sentiment analysis, based on Liu (2006). Second, we will perform a panel data model, to determine in which way online UGC is related to the sales of video games. When the volume of UGC does influence sales positively, a strategic decision could be to push consumers to write their review, as Bol.com already does with their consumers by the ability to win a Bol.com-cheque when writing a review. In another way, when the sentiment of UGC of a specific video game is negative, firms can analyse UGC to determine categories in which the negative sentiment lies. For example, graphics of a game, or just the slow delivery time of the firm were the product is purchased.

Studies that already generated some interesting results in a different industry are, for example, Chintagunta et al. (2010), who find that the valence (mean user rating) of online Word-of-Mouth has a significant impact on box office earnings for movies. Also, Liu (2006), found interesting facts on the volume on online UGC. They presented that when

the volume of online UGC is positively correlated to sales, companies can push buyers of a game to place UGC to stimulate their sales, or the other way around.
The time interval we will use to study these aspects are based on Eliashberg and Shugan (1997), who show that the first week for a movie determines the success of a movie. Research on the movie industry is important for our study, because the process of the release of a movie is somewhat similar to the release of a game. The success of a movie depends on the expectations before release, the opinion of experts and users, and the popularity of the genre. The same holds for the release of a videogame. In their research, Liu (2006) study the effect of online UGC as well, where they use the first 8 weeks after the release of a game. As we will extend their research, we will use the first 8 weeks after release as well and use panel data models to study the effect of UGC on sales over time.

1.1 Problem Statement and Research Objective

Sales are an important part of every company, also for companies that are active in the video gaming industry. Something affecting sales, according to Chintagunta et al. (2010) among others, is the online User-Generated Content, or UGC. The gaming community is very active online and likes to share their thoughts with each other. A question that may arise in User-Generated Content is to what extent consumers are influenced by this online UGC. For this reason, we will study the following research question:

RQ: How does online UGC affect the sales of video games in the first 8 weeks after its release?

We use the first 8 weeks because of the findings in Liu (2006), who perform a similar study for the effect of online UGC for movies on box office revenue. They find that most UGC is generated in the pre-release and opening week.

We answer our research question by using a panel data model, which estimates if and in which way the online UGC does influence sales. In this regression model, we define User-Generated Content with 4 attributes:

1. *Volume*: the amount of reviews.
2. *Negative sentiment*: the fraction of the negative reviews of the total amount of

reviews.

3. *Positive sentiment*: the fraction of the positive reviews of the total amount of reviews.
4. *Rating*: the rating that is given to a game.

The positive and negative sentiment of a review will be distinguished by using sentiment analysis. We will use a panel data model, because we are dealing with panel data, which are dependent on multiple variables:

1. *Time*: number of weeks after release
2. *ID:* the name of the game, in the model represented as an ID

We will discuss panel data models thoroughly in Section 4.

We will extract our UGC data from Amazon.com, one of the largest online retailers in the USA. Because real sales data of video games is not publicly available, we will use the educated guesses of sales for the USA from Vgchartz.com, which we will further explain in Section 3.3. We will only use data on video games released in 2014 for PlayStation 4, because players on different platforms have other preferences, which may lead to difference in online UGC, not depending on the games, as described in Zhu and Zhang (2010)

There are several papers that find that UGC does have a significant impact on consumer behaviour. For example, Chintagunta et al. (2010) find that the valence (mean user rating) of online UGC has a significant impact on box office earnings. Zhu and Zhang (2010) found that online reviews have more impact for games which are less popular and games whose players have greater internet experience. Because of these, and several other findings, our goal is to investigate which aspects of the reviews have a significant effect on the sales of a video game. In Liu (2006), they find that for movies, expectations before release are high, which may lead to a high rating before release. While in the opening week, people tend to be more critical. Tirunillai and Tellis (2012) find that User-Generated content has a large impact on the performance of a company on the stock market. With sentiment analysis, they classify reviews as positive chatter or negative chatter. They show that this User-Generated Content is correlated with abnormal returns and trading volume. They find that the volume of User-Generated Content has a strong influence. According to Flanagin and Metzger (2013), readers find the rating of a game with a lot of reviews more reliable than the rating of a game with less reviews. The rating is a very important part of an online review, and thus of the UGC. Chevalier and Mayzlin (2006) show that rating does affect the sales of books. They studied the reviews of the books on Amazon.com, because Amazon.com is the largest online store in the United States with a large number of reviews. In their research, they found that a 1 star review has more impact on sales than a 5 star review.

We will combine these papers to answer our research question on the influence of online UGC on video games. Because the sentiment of the online UGC is not directly available in our dataset we will use sentiment analysis based on Tirunillai and Tellis (2012) to classify the sentiment.

A detailed description of all measures used in this research will be provided in Section 3, the methods we will use are fully described in Section 4.

1.2 Scientific and Managerial Relevance

This study will be very relevant for companies that develop new products. Chintagunta et al. (2010) present that the effect of the rating and reviews on sales is significant. The rating of a game can be assigned in different ways. On Metacritic.com for example, a rating can be assigned in whole numbers from 1 to 10, where on Amazon.com, a star-rating can be assigned from one-star to a five-star rating. The review with this rating is the text accompanying the rating. The review mostly explains what the writer finds most attractive in a game or, when the rating is low, disappointing. The presence of an effect of these two on sales means that it is of great importance to understand how this User-Generated Content affects sales. Sales drive a company, because it ensures it to grow and exist. With the growth of the volume of User-Generated Content (UGC), it is therefore of great importance to understand in what way this influences sales. When we learn how UGC does affect sales, we can learn from it and use it in our advantage. Managers will be able to act on these findings by for example creating a team that reacts on these reviews and developers by taking issues, discussed in these reviews, into account in developing a new release of the game.

We will use online UGC because in the current society, where social media keeps growing, online content grows in importance as well. There are already studies that show the impact of UGC on purchase behavior of consumers. For example Chevalier and Mayzlin (2006) show that the reviews, ratings and the volume do affect book sales. Tirunillai and Tellis (2012) show the use of sentiment analysis to measure the effect of the online UGC on the stock market. The contribution we will have to this research is that we, first, use sentiment analysis to assign whether a review is positive of negative. Second, study in what way reviews of this week, affect the sales of the next week over a selected period of time for the video gaming industry, for which not much similar research has been done.

1.3 Structure of the Thesis

The structure of this research is as follows. In Section 2 we will explain our literature review and state our hypotheses. In Section 3 we present the data and general statistics. We also mention how we obtained and processed the data. In Section 4 we describe our methodology, we explain which models will be used and how we use them. In Section 5 we discuss and analyse the results. In Section 6 we will present the conclusion of our paper and topics for futher research.

2 Theory and Hypotheses

2.1 Literature review

In this section, existing literature with respect to our research question will be reviewed. We will provide our addition to existing literature in this section as well. For a better overview, we divided this section in to two parts: in Section 2.1.1, literature on the definition of UGC will be discussed and in Section 2.1.2 we will discuss literature on different methodologies for classification.

2.1.1 Defining UGC

Flanagin and Metzger (2013) already provide a method for analyzing people's sentiment towards movie reviews, in which they focus on the way in which a review is trustworthy. First, they study in what way the perception of the quality of information is influenced

by the source of social information. Second, they investigate the interaction between the source of the rating and volume of social information on the quality of the perception of the quality of information as well. To study these two topics, they use a random sample of 1,207 U.S.-adults that represented the population of the U.S. Each participant was presented a screen shot of one page of review, with the movie rating, source of rating and the volume of supplied ratings. Results indicated that the volume of ratings was positively correlated with trust and confidence in content. They found that opinions converged with the existing ratings they were exposed to. When there were just a few user generated reviews and also just a few expert reviews, the expert opinions were found more trustworthy. This paper shows that there are differences in the way in which something is presented to an individual. A review that is exactly the same with the name of a critic with it, can be interpreted more trustworthy than the same review without the name. For this reason, we will choose to obtain our UGC data from one source, to avoid this type of differences. It is important to know what attributes have an effect on the rating as well. In their research, they present their whole dataset to a panel, which we will try to automate due to time limitations.
Liu (2006), on the other hand, uses Word-of-Mouth (WOM) information for movies to study patterns of WOM in sentiment based on whether WOM is positive or negative. They observe that there is more WOM during pre-release and the openings week and that expectations before release are high, but reviews become less positive during the opening week. In their methods, they choose three judges to classify their WOM as positive or negative. We will partly use their methods for classifying as will be discussed in detail in Section 4.1.

Eliashberg and Shugan (1997) use classification of a magazine to assign sentiment to a review. They investigate whether reviews of critics influence the market performance or are predictors. In their paper, they find that reviews of critics do correlate with late and cumulative box office receipts, but that this correlation is not significant with early box office receipts. These findings are both for negative reviews with a negative effect on box office receipts as for positive reviews with a positive effect. They also find that reviews are more influencers than predictors, as far as their empirical findings suggest. This means that they influence the potential customers and that they are not predictors of the succes of a video game. Because we focus on in what way UGC affects sales, we will not only depend on the way Amazon.com assigned sentiment to reviews, but will perform a sentiment analysis in addition to our independent variables.

Where previous papers focused on the sentiment of reviews in their research, Chevalier and Mayzlin (2006) focus on other aspects of UGC. In their research, Chevalier and Mayzlin (2006) study the effect of book reviews on sales of books on two websites. They find that reviews were positive on both websites, but there are more and longer reviews on the more popular website. With the use of a regression model, they obtained that relative sales of the two websites are related to differences in the volume of book reviews and rating. This review shows that there is an effect of UGC on the sales of books, which leads us to think that UGC could have impact on the sales of video games as well.
Chintagunta et al. (2010), define the sentiment of a game with their rating, which is a similar method to the research of Eliashberg and Shugan (1997), who base the sentiment of their data on the classification of a magazine. Chintagunta et al. (2010) measure the impact of national online user reviews, represented by valence, or mean-user rating, volume and variance on box office performance of movies. In their research, they account for three complications with respect to analysis: (1) spatial aggregation, (2) serial correlation caused by sequential release of movies and (3) serial correlation caused by other unobserved components. They find that valence of online WOM has a significant and positive impact on box office earnings.

Paper	Industry	Estimation method	Number of observations	Main insights
Flanagin and Metzger (2013)	Movie reviews	- Statistic analysis	1,207	There are differences in reviews of experts and the public. Also credibility is different.
Liu (2006)	Movie reviews	- Statistic analysis - Regression	40 movies, 12,136 messages	Most WOM in pre-release and opening week. High expectation before release. Significant exploratory power WOM for aggregate weekly box office revenue.
Eliashberg and Shugan (1997)	Movie reviews	- Regression - Data analysis	2,104 reviews for 172 movies	Positive and negative percentages of the statistical predictors of the cumulative box office for all weeks after release.
Chevalier and Mayzlin (2006)	Book sales	- Regression - Cross sectional analysis	2,387 observations	Positive reviews, but longer and more on Amazon.com. Improvement in reviews $\rightarrow$ increase relative sales. 1 star rating greater impact than 5 star.
Chintagunta et al. (2010)	Movie industry	Regression, Generalized method of moments	Daily box office + reviews on 148 games during 16 months	valence of online WOM has a significant and positive impact on box office earnings.

Table 1: Literature overview UGC

2.1.2 Methods for classification

Tirunillai and Tellis (2012) use the Naive Bayes and Support Vector Machine classifiers, which are proven to be reliable for text classification applications, to identify whether a review is positive or negative. The reviews where the two classifiers are in agreement have been assigned to the classification. They aggregate the individual consumer reviews of a firm in any given day to obtain time series data. We will use this research, as we will first, create a training set of classified reviews based on the method of Liu (2006), and than create a Naive Bayes model to classify our other datasets, based on Tirunillai and Tellis (2012). We will combine these methods due to our time limitations on the other hand, and to analyse to what extent our sentiment variable corresponds to the sentiment given by Amazon.com, which is based on the rating.

Zhu and Zhang (2010) study how characteristics of products and consumers influence the perception on reviews with respect to buying video games. In their paper they find that reviews are more influential for games which are not that popular and for games with players that are not foreign with using the internet. This research shows us how to analyse the gaming reviews. For example the differences between the characteristics of consumers. They do, however, use the volume and rating of the review, as are some papers in Section 2.1.1. We will, therefor, extend their research by adding the methods of Liu (2006) and Tirunillai and Tellis (2012) to classify the sentiment of the games ourselves.

Lee and Bradlow (2011) show a method to support the analysis and visualization of market structure. They do this by automatically eliciting product attributes and brand's relative positions from online reviews. They make a new combination of existing text mining and marketing methods to describe an automated process for identifying and analyzing online customer reviews. This process is easily repeatable both for online reviews of physical products as for online reviews of services. The example they give is based on reviews of digital cameras. This article is important for our research because it shows how to use a good combination between text mining and marketing methods. From this article, we will use the methods for preprocessing before including our dataset in the Naive Bayes classifier.

Another way of studying UGC is done by Decker and Trusov (2010), who want to estimate aggregate consumer preferences using online product reviews. The main question is how to turn the available plentitude of individual consumer opinions into aggregate consumer preferences. In their study they used reviews from the cellphone market. The authors presented a complex econometric framework to turn the plentitude of individual consumer opinions into aggregate consumer preferences based on the online product reviews. During this process they made use of the conjoint analysis and negative binominal regression. Thus, this paper shows us a way to aggregate the customer reviews and what happens along the way. Klucharev et al. (2008) study human behavior in a more neural way. These other applications of human behavior explain why this subject is so important in classifying sentiment in more ways, because human behavior always keeps being unpredictable.

Paper	Industry	Estimation method	Number of observations	Main insights
Decker and Trusov (2010)	Mobile phones	Conjoint analysis - Negative binomial regression	20,000 reviews	The review-based results compare in a good way with consumer preferences obtained through conjoint analysis techniques.
Zhu and Zhang (2010)	Game reviews	- Two-stage nested logit model - Regression	220 games, 3,330 messages for PS2. 3,305 messages for XBOX	Online reviews are more influential for less popular and online games.
Tirunillai and Tellis (2012)	Stock Market	Sentiment Analysis, Vector Auto Regressive Model	Daily UGC of 4.5 years	UGC is correlated with abnormal returns, risk, and trading volume. Volume of chatter shows strongest relationship with abnormal returns and trading volume.
Lee and Bradlow (2011)	Reviews digital camaras	- Conjoint regression	8,226 reviews for 575 products	They described an algorithm for analyzing online customer reviews and presented a combination of existing text-mining and marketing methodologies.
Tirunillai and Tellis (2014)	Brand analysis	Latent Dirichlet Allocation	Identical to Tirunillai and Tellis (2012)	For vertically differentiated markets objective dimensions dominate and are similar over these markets. Low heterogeneity across dimensions, high stability over time. For horizontally differentiated markets subjective dimensions more dominant but vary over time. High heterogeneity across dimensions, not stable.

Table 2: Literature overview methods classification

2.2 Research approach

The main purpose of this paper is to study the effect of online UGC on sales. To answer the research question we have to collect a lot of data, which we will collect using a Web Crawler. A detailed description of the use of this Web Crawler and the process of obtaining data will be provided in Section 3.4. We will extract data on sales and UGC from different websites, which we will describe in Section 3. If we obtained all the data, we process the online UGC and the sales data with RapidMiner. We extract the sales data per week per game for the first 8 weeks after release. The same applies for the UGC, we extract the text based review and rating per week per game. We will use RapidMiner to perform a sentiment analysis as well. Through this way we can divide the data into three sentiment groups: (1) positive, (2) negative and (3) neutral. After all data is processed, we will divide our data into dependent variables (*Sales*) and independent variables (*Rating*, *Volume*, *Positive sentiment* and *Negative sentiment*) remain. With these variables we will perform panel data models. As a result, we might be able to observe which variables have a significant impact on sales.

2.3 Dependent Variable

The dependent variable in this research will be the sales of video games. The sales numbers are more often used as dependent variable. For example Chevalier and Mayzlin (2006) use as dependent variable in their research on book reviews and Eliashberg and Shugan (1997) use sales as dependent variable when studying critics' reviews. This shows that sales is important in current literature. After all it is crucial for these companies to have sales.

In our research we will try to explain changes of sales by online UGC. All independent variables we use to explain the sales will be extracted from online UGC.

2.4 Key Independent Variable(s) and their Relation with the Dependent Variable(s)

The key independent variables we will use in this research will be extracted from online UGC. Tirunillai and Tellis (2012) use online UGC to extract their independent variables when studying the relationship between UGC and the stock market.

Our independent variables are based on Tirunillai and Tellis (2012), who use similar variables in their research. The independent variables we will use are the following:
Rating: One of the measurements of UGC is the rating given with a review. These ratings are on a numerical scale from one to five. Here one is the most negative rating and five the most positive. We use the mean rating per video game per week.
Volume: Volume describes the total number of the reviews (UGC) for a specific video game in a specific week.
Negative sentiment: Negative sentiment refers to the reviews with a negative sentiment as a percentage of the volume of a certain week for a certain game. In Liu (2006) they also use the percentage of volume to determine sentiment. We will define this sentiment by classifing the reviews by using the Naive Bayes classifier, which will be discussed in Section 4.
Positive sentiment: Positive sentiment refers to the reviews with a positive sentiment as a percentage of the volume of a certain week for a certain game. We will define this sentiment by classifing the reviews by using the Naive Bayes classifier, which will be discussed in Section 4.

As we mentioned earlier we are dealing with panel data, which need to be approached in a different way because of the fact that they depend on two dimensions. The two variables influencing the other variables are:
Time: The variable time stands for the weeks after release. One time unit is one week.
ID: The ID stands for the video game. Each video game will get a number, which forms the ID of an individual video game.

We will study the relationship between the independent variables and the dependent variable *Sales*. With panel data models, we will investigate whether our independent variables do have a significant impact on the dependent variable and if so, what the impact is on the dependent variable.

3 Data

In this research, we will use data from two different sources. For our online UGC we will use Amazon.com, one of the largest online retailers in the USA and for our sales data we use Vgchartz.com. Vgchartz.com performs educated guesses on sales of video games in units. Their educated guesses of sales are on the markets of Europe, the USA, Japan and Global. For this research, we will focus on data of the USA, because this is in line with our online UGC data, which we extracted from Amazon.com, an online retailer in the USA. Amazon.com sells video games for all platforms, but as previously mentioned, our focus will be only on PlayStation 4.

The reason for this is because there are games available for different platforms, but these games differ in for example look and feel for these different consoles. Another reason is that players on different platforms may have other preferences for games, which leads to a difference in review and rating. On Amazon.com, the same games on a different console are treated like separate games, where customers will write their review on the correct game for the correct platform. For our reseach we only use action games launched in 2014 and the reviews of the first 8 weeks after release. On Vgchartz.com the sales data is available for all the games. In Figure 1 we show how we process the data.

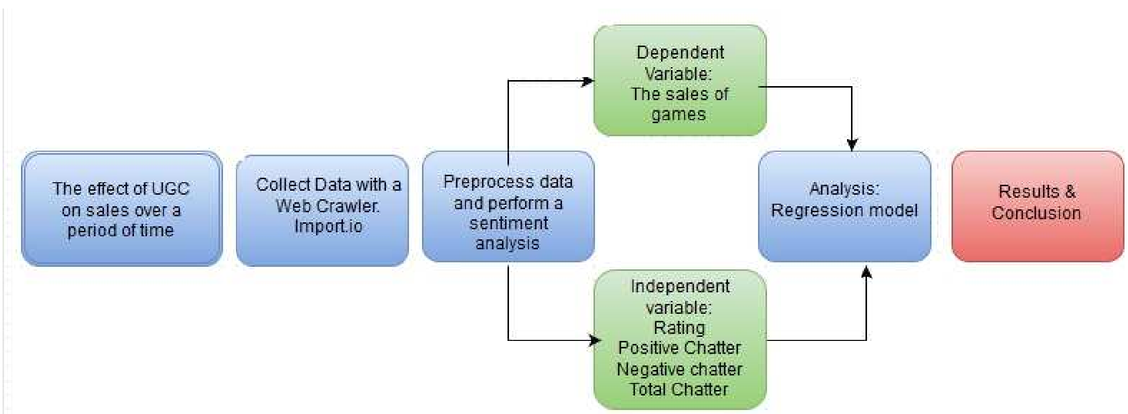


Figure 1: Flowchart of the data flow used in this research.

First we collect all data of the different websites. For this process we use a Web Crawler, described in Appendix B. After we collected all data, we will preprocess the data so we only have reviews and sales data of the first 8 weeks after the release of the games. While preprocessing the data, we will calculate in which week after release the review is written by subtracting the release date from the date at which the review is written. After preprocessing the data, we will perform our sentiment analysis, described in Section 4.1.

3.1 Empirical settings: the video game industry

The video game industry already exists since 1972, when Magnavox introduced Odyssey, the first home video game console. There were 12 games which could be played on the console and over 100,000 consoles were sold by the end of the first year. Now, after already seven generations of game consoles are released, a lot has been changed. In the beginning of the video game industry, games and consoles were usually produced by the same firm, but now, publishers, like Electronic Arts, have been specialized in a way that they are able to create the same game for different consoles. The popularity of video games has been changed as well. Already in 2001, playing video games was rated as the top entertainment when competing with renting movies, watching tv, surfing on the internet and reading books IDSA 2001. By 2010, 55 million seventh generation consoles were sold only in the U.S. by March 2010, according to Digital Digest 2010.

The structure of the video game industry, according to Kaiser (2002), is a prototypical platform market, where the term prototypical refers to 'having the typical qualities of a particular group or kind of person or thing' according to the English dictionary. In this market, the console acts as a platform with two different end users, with different interests: (1) the end user, or consumer of the console and (2) the game developers. There are, thus, three classes of players in the market, which are all related as in Figure 2.

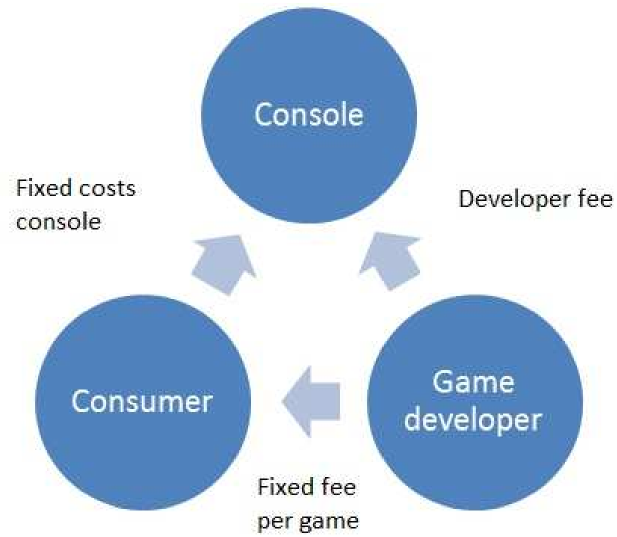


Figure 2: Video game industry

A consumer buys a console to buy a video game at a fixed price for the console and a fixed price for the video game. In order for the consumer to play a game, the game developer has to pay a fee in order for him to receive the rights of the code that allows him to make his video game compatible with the console. This fee is not a fixed fee, but has to be paid for each copy of the video game that is sold. There are two types of video games: first party video games and second party video games, where the first are video games produced by the producing firm of the console and the second are firms, independent of the manufacturer of the console, that make the video game accessible for all consoles, due to high fixed costs for production.

According to Statista.com, most consoles sold in 2014 were PlayStation 4, with almost 14 million unit sales, globally. They sold over 30% of all consoles in that year. According to the same source, revenue on video games for consoles exceeded mobile video games and PC video games for 2012 and 2013. The top genre for video game sales in 2014 was 'action', with almost 29% sales in the U.S. in 2014 followed by 'shooter' with almost 22% sales. These statistics support our decisions for analyzing data on the PlayStation 4 in 2014 for action games.

3.2 Online UGC data

For this research, dat on online UGC is required which is explained by the key variables described in Section 2.4. Our online UGC must consist of a textual part, where a review is written, a rating can be assigned and the date of publishing is present. We would like our data to be representative of the average online consumer, which is the reason why we use data from Amazon.com to approximate our online UGC. Amazon.com is one of the largest online retailers of the USA. They offer a variety of products, from clothing to car parts, to electronics and video games as well. The games from which we will collect our data are all action video games released in 2014.

A preview of a game for a potential customer is presented in Figure 3. In this figure, we observe that we are immediately drawn to the stars that represent the rating. With these stars, a customer is able to view how many reviews there are written and by clicking on it, he or she is directly navigated to the reviews.

Figure 3: Main data of a game on Amazon.com

Figure 4 represents the way reviews are presented. First, an overview is displayed with the fraction of customers divided over the ratings. Second, the reviews are presented in a list, where there are 10 reviews per page. Potential customers are able to sort on most helpful or most recent and are able to filter on customers who bought the game on

Amazon.com, on the rating and on the platform.

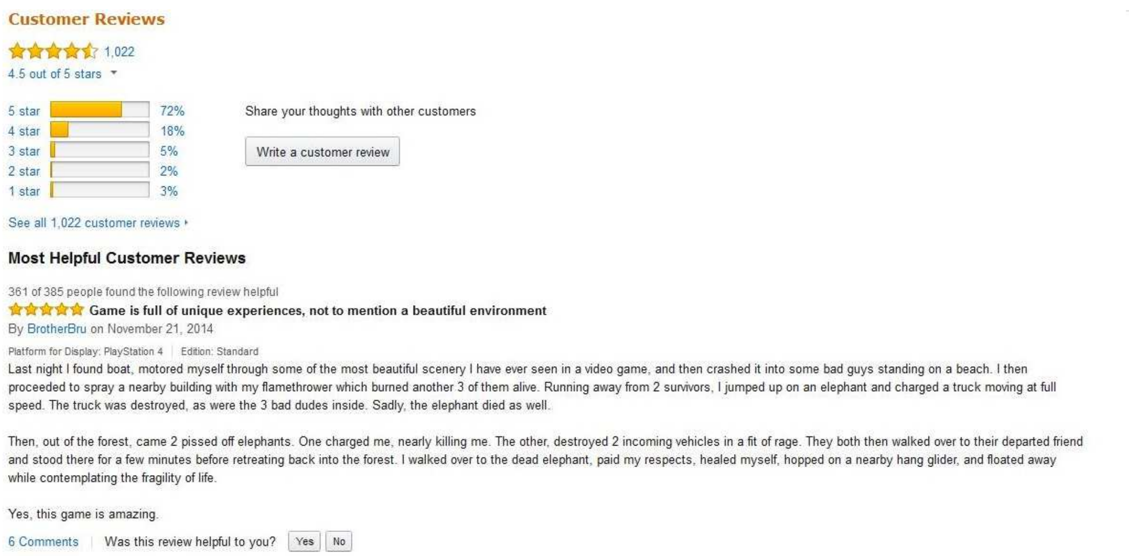


Figure 4: An overview of the reviews.

We obtained data for 37 action games released on the PS4 in 2014, with a total of 17,632 reviews divided over 1,772 pages where we obtained the text of the review, the rating and the date of posting.

3.3 Sales data

For our sales data we will use data of Vgchartz.com. On this website, educated guesses of the sales in units of video games are available. We will use these educated guesses, because real sales data of video games is not available on the internet and because our goal is to study the effect of online UGC on sales, these educated guesses will suffice.

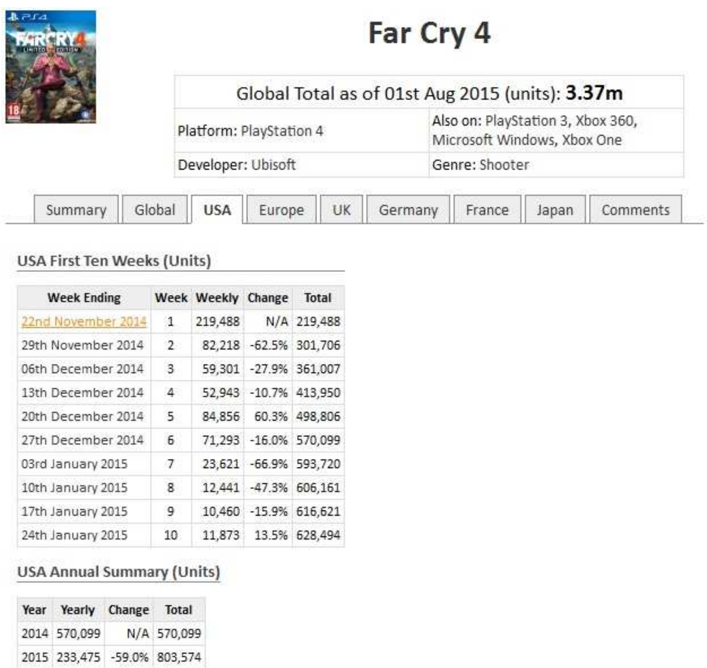
According to Vgchartz.com the method with which they create their estimates is based on the following procedure[1]:

- *Passively polling end users to find out what games they are currently purchasing and playing*
- *Polling retail partners to find out what games and hardware they are selling*
- *Using statistical trend fitting and historical data for similar games*

[1] Vgchartz.com/methodology.php

- *Studying resell prices to determine consumer demand and inventory levels*
- *Consulting with publishers and manufacturers to find out how many units they are introducing into the channel*

They also state that '*All data is regularly checked against manufacturer shipments and data released publicly from other tracking firms to ensure accuracy.*' and that their methods are 'ever-developing'.


Far Cry 4

Global Total as of 01st Aug 2015 (units): **3.37m**	
Platform: PlayStation 4	Also on: PlayStation 3, Xbox 360, Microsoft Windows, Xbox One
Developer: Ubisoft	Genre: Shooter

Summary | Global | USA | Europe | UK | Germany | France | Japan | Comments

USA First Ten Weeks (Units)

Week Ending	Week	Weekly	Change	Total
22nd November 2014	1	219,488	N/A	219,488
29th November 2014	2	82,218	-62.5%	301,706
06th December 2014	3	59,301	-27.9%	361,007
13th December 2014	4	52,943	-10.7%	413,950
20th December 2014	5	84,856	60.3%	498,806
27th December 2014	6	71,293	-16.0%	570,099
03rd January 2015	7	23,621	-66.9%	593,720
10th January 2015	8	12,441	-47.3%	606,161
17th January 2015	9	10,460	-15.9%	616,621
24th January 2015	10	11,873	13.5%	628,494

USA Annual Summary (Units)

Year	Yearly	Change	Total
2014	570,099	N/A	570,099
2015	233,475	-59.0%	803,574

Figure 5: Representation of the data on sales on Vgchartz.com

Figure 5 shows the view of the sales numbers for a video game. In this figure, we observe that weekly sales numbers are available in the first 8 weeks after release for different regions and global numbers as well. We will use sales data for the USA because this is consistent withour UGC data from Amazon.com. In our dataset, we collect the release date, weekly totals and weekly sales in units. An annual summary is presented as well, but since our goal is to study the effect of online UGC on sales per week, we will not use this data.

3.4 Data Collection

To collect all data, we use a Web Crawler. A Web Crawler is an internet bot that extracts data from web pages in a systematic way. When using a Web Crawler, a website will be selected and the indicated parts of the pages will be extracted. The data we collect, is collected with Import.io. The data we want to recover from the Web Crawler are of two types: (1) User-Generated Content and (2) Sales. We will extract our UGC-content of the website Amazon.com, one of the biggest online retailers in the USA. Because actual sales records of videogames are not easy to obtain, we will use the educated guesses of sales provided by Vgchartz.com for the USA. We selected 37 games of the same genre all released in 2014. We want to obtain data on ratings of the user, the date at which the review is written and the review itself. All collected data will be processed with RapidMiner, a software platform that provides an integrated environment for data mining and machine learning, among others. Text mining is also possible with RapidMiner but you have to install a plug-in. We will preprocess the data with RapidMiner so we can use data for our sentiment- and regression analysis. A detailed description of the data extraction is presented in Appendix B.

3.5 Definition of Measures

As is already discussed in the previous sections, we will use measures from different sources. For our sentiment analysis, we need three variables: (1) *Volume*, (2) *Negative sentiment* (3), *Positive sentiment* and (4) *Rating*. By using sentiment analysis, we will classify all reviews as positive, negative or neutral which leads to the postive and negative sentiment, respectively. We will use the rating to complete the extracted variables from the UGC data.

For our panel data model, we will use the natural logaritm of sales. The reason for the log specification is based on Chevalier and Mayzlin (2006), who state that the log-specification estimates the effect of a change in the independent variables on the percentage change in the dependent variable. *Volume* will be the number of reviews depending on the game and time, which is easier to use. The variables *Negative sentiment* and

Positive sentiment are expressed as a percentage of the total volume and *Rating* is an ordinal variable. All the variables are dependent of the game *ID* and *Time*.

No.	Variable name	Description	Measurement	Type
1	*Sales*	Sales of a game per week in units	Units	Numeric
2	*Time*	Week after release from t=1,...8	Scale	Ordinal
3	*ID*	ID number of a game	Scale	Nominal
4	*Volume*	Total amount of UGC per week per game	Scale	Numeric
5	*Negative sentiment*	Percentage of negative sentiment per week per game	Percentage	Numeric
6	*Positive sentiment*	Percentage of positive sentiment per week per game	Percentage	Numeric
7	*RATING*	Average rating of a game per week	1,...,5, with 1 the worst and 5 the best	Ordinal

Table 3: All measures used in this research

3.6 Data Description

We collected data both from Amazon.com and from Vgchartz.com for 37 games released in 2014. In Appendix A all games we used in this research are presented together with their descriptive statistics. Table 23 shows total sales, the amount of reviews and the average rating given by the reviewers for each game.

3.6.1 Sales data

Figure 6 shows that most units of the video games are sold in the first week after their release. In the second week the number of units sold drops a lot compared to Week 1.

In the third week there is another drop in sales but it is less than the drop in Week 2. The left figure in Figure 6 shows that after Week 3 the sales stabilizes and that sales are somewhat stable over the remainder of the weeks.

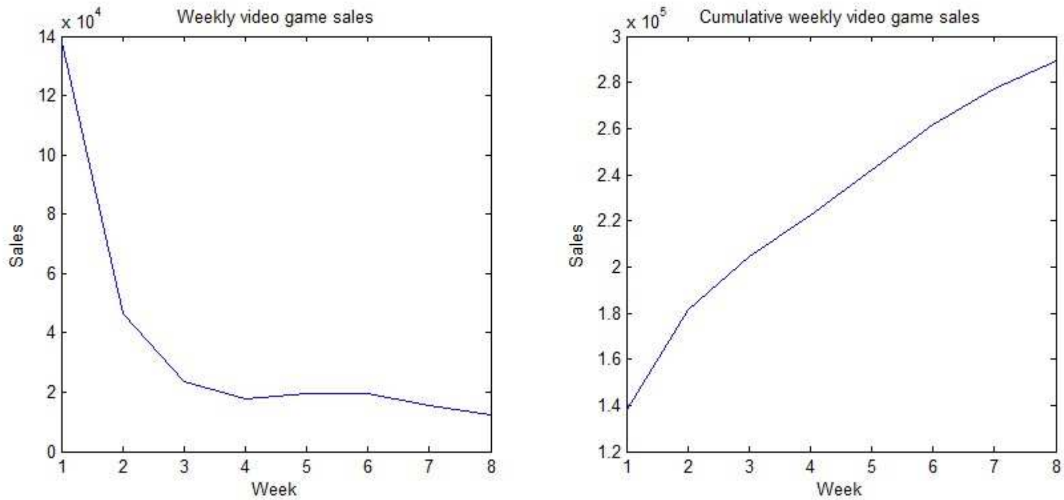


Figure 6: The average sales per week and graph two shows the cumulative sales of the games

In Table 4 we observe that standard deviations are large. This can be explained by the fact that we didn't select the games based on sales, which causes sales numbers to vary a lot for different games. When we take a look at Table 23 in Appendix A, for example, we observe that sales for 'Grand Theft Auto V' are much larger (1,709,222 units) than for 'The Lego Movie Videogame' (51,430 units). The minimum and maximum confirm the big differences, because these indicate a large difference as well. Again, Table 23 shows these differences, with the best game sold in the first 8 weeks is 'Call of Duty: Advanced Warfare' with 1,878,115 units sold and the worst selling game is 'Sherlock Holmes: Crimes and Punishments' with only 4,388 units sold.

Week	Mean	Std Deviation	Median	Min	Max	total weekly sales
1	137,893.75	220,562,67	34,564	956	874,427	5,515,750
2	46,650.49	82,615.20	16,473	475	477,358	1,912,670
3	23,477.68	35,695.02	8,102	227	176,985	962,585
4	17,862.63	26,255.53	4,721	154	105,166	732,368
5	19,723,29	36,169.88	4,177	124	179,466	808,655
6	19,417.24	41,551.22	3,794	133	225,354	796,107
7	15,344.15	36,225.51	3,535	131	205,565	629,110
8	12,143.46	32,444.54	3,385	105	207,710	497,882

Table 4: Statistics of the weekly sales over the first 8 weeks after release

We use the the natural logaritm of sales, because, usually, sales are exponential. As we observe in Figure 7 in comparison to Figure 6, the logaritm of sales is much more linear, which we are able to use in our panel data models.

Week	Mean	Standard Deviation	Median	Minimum	Maximum
1	10.650	1.721	10.444	6.863	13.681
2	9.579	1.729	9.707	6.163	13.076
3	8.943	1.694	9.000	5.425	12.084
4	8.705	1.653	8.460	5.037	11.563
5	8.576	1.726	8.337	4.820	12.098
6	8.427	1.760	8.241	4.890	12.325
7	8.208	1.747	8.170	4.875	12.234
8	8.196	1.565	8.127	4.654	12.244

Table 5: Descriptive statistics of the weekly log(sales) over the first 8 weeks after release

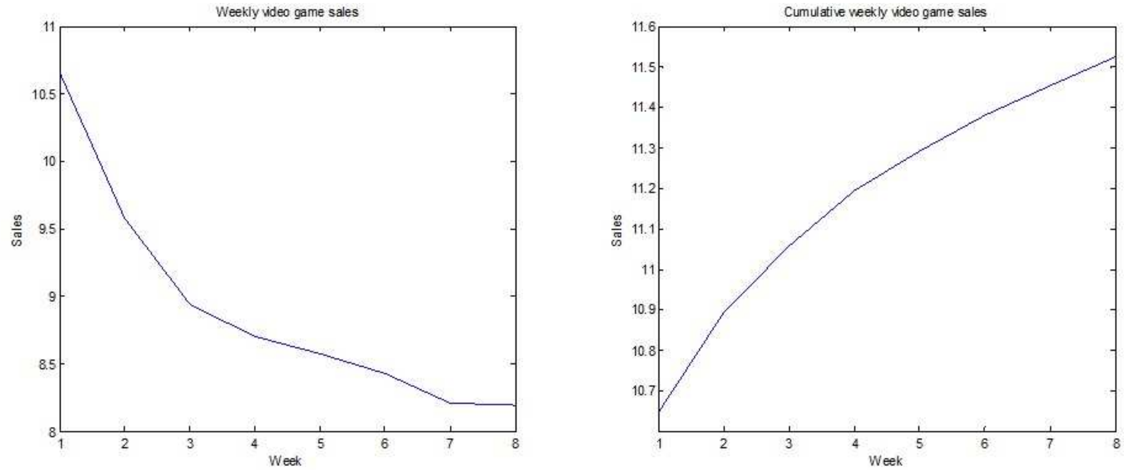


Figure 7: The average log(sales) per week and graph two shows the cumulative log(sales) of the games

When we look at Table 6 we observe the descriptive statistics based on the games. We see that there are large differences between the number of reviews per game. The game 'Destiny' has the most reviews in the first 8 weeks, with 615 reviews. The 'Game Dynasty Warriors 8: Xtreme legends' has only 1 review in the first 8 weeks. When we look to the means of the rating per game, we observe that the mean is 3.94. This means that the people generally are positive about the games they write reviews about. The standard deviation shows us that almost every game has a value above 1 and below 2 with a mean of 1.445. On a scale from 1 to 5, this shows that the rater does not always agree. The minimum and maximum confirm this, the most games have a minimum of 1 and a maximum of 5.

Game	Mean Rating	Min Rating	Max Rating	Std. Rating	Count Reviews
Alien: Isolation	4,02	1	5	1.394	104
Assassin's Creed: Unity	3.02	1	5	1.66	450
Bound By Flame	3.2	1	5	1.288	60
Call of Duty: Advanced Warfare	3.6	1	5	1.576	230
Destiny	3.45	1	5	1.549	615
Dragon Age: Inquisition	4.41	1	5	1.103	242
Dynasty Warriors 8: Xtreme Legends	5	5	5	0	1
Far Cry 4	4.5	1	5	1.003	214
Final Fantasy XIV: A Realm Reborn	4.03	1	5	1.527	60
Grand Theft Auto V	4.48	1	5	1.115	284
Guilty Gear Xrd: Sign	4.5	1	5	1.067	18
inFAMOUS: Second Son	4.45	1	5	0.945	31
Just Dance 2015	3.25	1	5	1.699	20
Lara Croft and the Temple of Osiris	3	1	5	1.604	7
LEGO The Hobbit	4.5	3	5	0.806	10
LittleBigPlanet 3	3.82	1	5	1.266	11
Lords Of The Fallen	3.73	1	5	1.47	55
Metal Gear Solid: Ground Zeroes	3.52	1	5	1.57	252
Middle Earth: Shadow of Mordor	4.66	1	5	0.887	267
Minecraft	4.85	1	5	0.755	27
murdered soul suspect	3.86	1	5	1.125	42
Plants vs Zombies: Garden Warfare	3.25	1	5	1.785	44
Rayman Legends	4.83	3	5	0.446	35
Samurai Warriors 4	4.09	2	5	0.9	11
Sherlock Holmes: Crimes & Punishments	4.27	2	5	1.052	11
sniper elite 3	3.97	1	5	1.257	31
The Amazing Spider-Man 2	3.26	1	5	1.421	46
The Evil Within	3.58	1	5	1.541	182
The Last of Us	4.77	1	5	0.772	524
The LEGO Movie Videogame	4.48	2	5	0.906	21
The Walking Dead: Season Two	4.16	1	5	1.225	19
The Wolf Among Us	4.93	4	5	0.249	15
Thief	3.28	1	5	1.442	134
Tomb Raider	4.71	1	5	0.736	107
Warriors Orochi 3	4.56	3	5	0.685	9
Watch Dogs	3.94	1	5	1.278	363
wolfstein the new order	4.37	1	5	1.11	114
WWE 2K15	2.85	1	5	1.596	95
Grand Total	3.94	1	5	1.445	4762

Table 6: Descriptive statistics of action games

3.6.2 UGC data

From the website Amazon.com we obtained the UGC data, where the most important data are the rating and the reviews. From Table 8, we observe that the rating most people assign to a video game is a 5 star rating, the highest rating they are able to assign. After the 5-star rating, the 4- and 1-star rating are mostly assigned, but there is a large difference between the amount of times a 5-star rating is assigned relative to all other ratings. Figure 8 below shows the distribution of the grades.

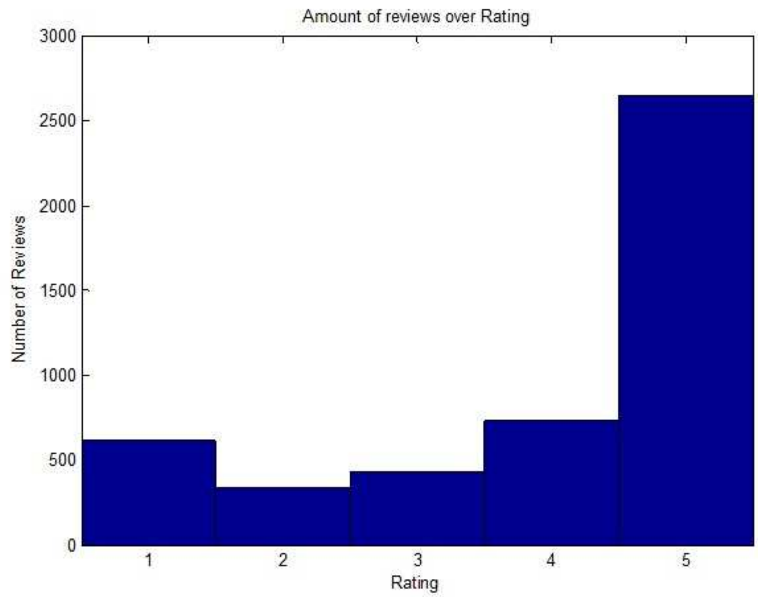


Figure 8: Histogram on the spread of the rating over all weeks

Table 7 shows that the average rating in the second week after release is the lowest. It also shows that when time passes the users assign better grades to the games. The standard deviation over time declines, which may indicate that users become more moderate in their way of evaluating the games. Table 7 also shows that in every week there are people rating the game best and worst, with a 5- and 1-star rating, respectively.

Week	Mean	Standard Deviation	Median	Minimum	Maximum
1	3.85	1.506	5	1	5
2	3.76	1.512	4.5	1	5
3	3.82	1.475	5	1	5
4	3.93	1.456	5	1	5
5	4.06	1.392	5	1	5
6	4.02	1.417	5	1	5
7	4.17	1.276	5	1	5
8	4.20	1.274	5	1	5

Table 7: Descriptive statistics of the rating over the first 8 weeks after release

Table 9 shows most reviews are written in the first week. After the first week, the amount of reviews declines quite fast. Beside week 1, week 2 has much more reviews than the other weeks. Table 8 show that the number of reviews in week 3 to 8 have about the same number of reviews. This may indicate that popularity declines fast after release of a video game, but becomes stable after week 3.

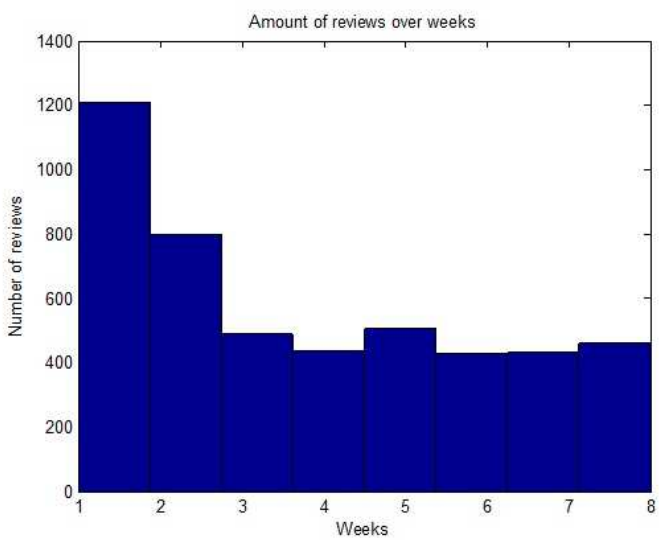


Figure 9: Histogram on the spread of the reviews over all weeks

	1	2	3	4	5	6	7	8	Total
no. Reviews	1208	800	490	438	506	427	432	461	4762
% Reviews	25.37	16.8	10.29	9.20	10.63	8.97	9.07	9.68	100.00

Table 8: Table of the spread of the reviews over all weeks

4 Research Methodology

In this section we will discuss the methodology that we will use in this paper. We will show what tests are needed to perform our panel data model and which model we will use. Through this way we want to provide a better insight in this study. We will use references to other articles to support our research methods.
In Table 9 we show the methods we use and the order we use them in.

Process	Method
Sentiment analysis	Naive Bayes
Panel data models	Fixed Effects model Between estimator Random effects model
Evaluation of the models	Hausman test Mundlak test

Table 9: Overview methodes used

4.1 Sentiment Analysis

For the sentiment analysis we will use the paper of Tirunillai and Tellis (2012) as framework. In their research Tirunillai and Tellis (2012) use the Naive Bayes classifier, which is proven to be reliable for text classification applications, to identify whether a review is positive or negative.

Figure 10: Process sentiment analyse

The process of this analysis is presented in Figure 10. We will first create a training set, where sentiment is assigned manually by three judges. Second, we will use this training set to create our Naive Bayes model, which we, third, perform on our test set. We, fourth, calculate the performance of our model and finally create the variables, used in our panel data model.

We will create our training set which consists out of 20% of our total dataset. We will draw this random sample from a uniform distribution, where each data point has the same probability of being in the training set. The reason for this is to avoid our training set from being unbiased. The test set will consist out of the whole data set, including the training set and consists out of 4762 rows.

The second part of the process is assigning sentiment to the training set. The method we will use for this assignment is based on Liu (2006), who invites judges to assign sentiment to all reviews. We will invite three judges who all have to read the reviews and assign sentiment to them. The three values they could assign to a model are negative (-1), neutral (0) and positive (1). To investigate the reliability of the readers, we will perform an intra-rater reliability test which scores how much homogeneity there is between our judges. The methods we will use to estimate this intra-rater reliability are the intra-class correlation coefficient and Cronbach's alpha, which are proven to be very reliable in several articles, for example by Towns (2013), Dixon et al. (2001) and You et al. (2015). Intra-class correlation describes the resemblance of units in the same group. The method assesses the consistency in which way the results are reproducible when assigning the sentiment by different observers for the same data. Cronbach's alpha, however, tests whether several items in a group are allowed to form that certain group and, thus, validate the creation of our three groups. Cronbach's alpha only takes into account the correlations within the groups and is an estimation for the reliability within the group. It does not take into account the validity of the groups. By combining both measures, we do take into account the validity of the groups and the reliability within the group. When the judges are proven to be reliable, we will create the final sentiment of our training set by assigning the value of the sentiment of which at least two judges are in agreement. When no judges are in agreement, the observation will be deleted from our used training set. We will check the robustness of the estimate by comparing our sentiment value with the rating assigned by the person who wrote the review. Here, a rating of 1 or 2 is assumed to be negative (-1), a rating of 3 will be neutral (0) and 4 and 5 are assumed being positive (1). We will compare this value by calculating the hit rate, which is defined in the following way:

$$Hitrate = \frac{\sum \#samevalue}{\sum Observations} \tag{1}$$

Where $\#samevalue$ are the number of items of which the judges are in agreement with the rating.

After determining the sentiment of our training set, we will create the Naive Bayes classifier, to assign sentiment to our test set. The process in RapidMiner, which is fully described in Appendix C and D, starts with loading our training set. We will select the attributes we need, to perform our model, which are the ID of the data, the review and the sentiment, which we assigned in the previous step of Figure 10. Before creating the model we, first, need to preprocess the reviews in order to provide a consistent dataset, which can more easily be analysed by our model. We preprocess our reviews, based on Lee and Bradlow (2011). First, we will tokenize the reviews, which segments the reviews into, for example, words, phrases, symbols. Second, we will convert all letters to low case, so RapidMiner does not distinguish the same words with a different combination of upper and lower case. Third, we will filter our tokens on text length, which will try to exclude non-existing words like "amazbalzzzzzzzzzzzzzzzzzzz" or "Argggggggggggggghhh", which were observed in our dataset. Finally, we will remove stopwords as 'a', 'is' or 'you', which are no relevant additions to the content of the reviews. Now the data is preprocessed we are able to create the Naive Bayes classifier.

According to Russell and Norvig (1995), Naive Bayes is a technique for constructing classifiers. A Naive Bayes classifier is a model that assigns to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. Naive Bayes is very popular in text classification applications. For example, it is used by spam filters for emails and even to classify large scale web-pages. Our approach is similar to Tirunillai and Tellis (2012). The main aim of their, and our, research is to classify the reviews into categories, which are positive sentiment, negative sentiment and neutral sentiment. The Naive Bayes classifier uses the training set to indicate classes and tests the model by calculating the likelihood of the review belonging to all classes for that review, the review is put into the class with the highest likelihood. In our case this method will give an output, similar to the training set, which is a vector of the three values, determined in the previous step: negative (-1), neutral (0) and positive (1). RapidMiner optimizes the model by calculating the performance for 10 times. The model with the highest performance will be used as output. Here, the performance is calculated by the fraction the model assigns the same value as the training set, by performing the model to the training set, which is represented by the assigned value of the judges. We will calculate the overall performance for evaluating our model in the same way.

To use the sentiment in our panel data model, we will aggregate the individual consumer reviews for all games in the same way as Tirunillai and Tellis (2012) with the reviews per firm. where we create two variables for the positive and negative sentiment respectively, which are calculated by the percentage of positive and negative reviews over all reviews per week per game and negative reviews for the variables $Positive\ Sentiment_{it}$ and $Negative\ Sentiment_{it}$, respectively.

4.2 Panel data models

According to Janssens et al. (2008), panel data refers to '*observations at multiple (consecutive) points in time for several subjects*'. This is exaclty what characterizes our dataset, which means we will use panel data models to perform our analysis.

There are several panel data models, where the Fixed Effects model and Random Effects model are the two most common, according to Baltagi (2008). The big difference between the two models is the underlying assumption. The assumption for the Fixed Effects model indicates that the individual specific effect is correlated with the independent variables which, in our dataset, means that there is an individual specific effect for the different games. With Random Effects models, the underlying assumption implies that (some of) the independent variables will be treated as if they were from random causes. We will perform both models and will test their consistency with the Hausman test, as presented by Hausman (1978). We will then, analyse our results with the preferred panel data model. After performing the Hausman test, we will justify our results by using a Mundlak test after this, as presented in Mundlak (1978).

Before performing our analysis with both models we will, first, analyse different types of variances, related to the models. With this type of data there is not only the overall variance, but the within variance, related to the Fixed Effects model or *within* effects model, as well as between variance. Here, the within variance generates the variance between the games over all weeks and the between variance generates the variance within one game over several weeks. We will calculate all three variances in the following way:

The overall variance is calculated in equation (1) by calculating for every variable the difference between the value of the variable for each game (i) and for each week (t) and the average value of that variable. The square of this variable is summed over all weeks and all games and divided by the product of the total games (N) and total weeks (T) minus 1. The variance formula's ar based on Wooldridge (2010).

$$S_0^2 = \frac{1}{NT-1}\sum_i \sum_t (x_{it} - \overline{x})^2 \tag{2}$$

The between variance is calculated by the difference between the average value of the variable of the game and the average overall value of the game, for each game. Here, we will, thus, look at the variance between all games over *all* weeks.

$$S_B^2 = \frac{1}{N-1}\sum_i (\overline{x}_i - \overline{x})^2 \tag{3}$$

For the within variance, we will calculate the difference between the value of the variable for each week and each game and the average value of that variable per game over all weeks as is stated below.

$$S_W^2 = \frac{1}{NT-1}\sum_i \sum_t (x_{it} - \overline{x}_i)^2 \tag{4}$$

After performing and analysing these calculations, we will perform our panel data models by using the following equation:

$$y_{it} = \alpha + x_{it}\beta + \mu_i + v_{it} \tag{5}$$

Here, y_{it}, our dependent variable, is represented by a vector of our weekly sales data for all games. x_{it} will be defined as a matrix of our independent variables *Positive sentiment*, *Negative sentiment*, *Volume* and *Rating.* μ_i in this model is our time invariant individual effect, which will be a result of this analysis. For the Fixed Effects model, the assumption holds that the individual effects are correlated with our independent variables as presented below:

$$COV(X_{it}, \mu_i) \neq 0$$

According to Wooldridge (2010) we need to capture heterogeneity which is done by the model by calculating the OLS-estimator with the following formula:

$$\hat{\beta}_{FE} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{y} \tag{6}$$

where $\tilde{X} = X - \overline{X}$ and $\tilde{y} = y - \overline{y}$. This transformation calculates the variables as the deviation of the average of all individual groups. Individual effects are now calculated as $\hat{\mu} = \overline{y} - \overline{X}\beta$. For the Random Effects model, it holds that individual effects are assumed to be random. Here the model is rewritten as:

$$y_{it} = \alpha + x_{it}\beta + u_{it} \tag{7}$$

with $u_{it} = \mu_i + v_{it}$ and the individual effects (μ_i) and disturbances (v_{it}) independent of the independent variables which is in contrast to the Fixed Effects model. $COV(X_{it}, \mu_i) = 0$ and $COV(X_{it}, v_{it}) = 0$ for all i and t.

In addition to the Fixed Effects model, which focuses mainly on the within effect, we will perform a between estimation as well. Here, the average of all variables over the games is regressed against the average sales, where the estimator is calculated as follows:

$$\hat{\beta}_{BE} = (\overline{X}^T\overline{X})^{-1}\overline{X}^T\overline{y} \tag{8}$$

We will evaluate the fit of the models using the F-statistic. For comparing the Fixed Effects model and the Random Effects model, we will use the Hausman test which compares the estimators, $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$, respectively, which are consistent under H_0. Under H_1 only the Random Effects estimator is consistent. Thus, under H_0, the Fixed Effects estimator is preferred, whereas under the alternative the Random Effects estimator is preferred. Under H_0, the following holds:

$$H_0 : \hat{\beta}_{FE} - \hat{\beta}_{RE} = 0 \tag{9}$$

For justifying the results of the Hausman test, we will use the Mundlak test, which suggests the estimation of the regression below:

$$y_{it} = \alpha + X_{it}\beta + \overline{X}_i\gamma + \mu_i + v_{it} \tag{10}$$

With i *= 1,...,n,* t *= 1,...,*T_i*.*

Here, $\overline{X}_i$ are the variable group means. According to Mundlak (1978), the Fixed Effects model is a Random Effects model, but with random effects that correlate with the explanatory variables. If these random effects are not correlated with the explanatory variables, $\gamma = 0$ and the Random Effects model is being preferred, when $\gamma \neq 0$, the random effects are correlated with the explanatory variables and the Fixed Effects model is preferred.

5 Analysis and Results

In this section we will discuss our analysis and results. This section is split into two sections. In Section 5.1, we will discuss results of the sentiment analysis. We show our results and analysis and present how we use these results in our panel data model. In Section 5.3 we present the results of the panel data models we used. We show how we performed the model and discuss our results and analysis.

5.1 Sentiment Analysis

After we created the process in RapidMiner, we run our program. The training set we created is a randomly drawn set from our overall data, which contains 20% of this data.

This dataset contains 933 reviews with their sentiment. There were 19 reviews for which no judges were in agreement. These reviews were excluded from our final training set. Around 77% of the sentiment was assigned by unanimous decision and 21% by split decision. The remaining data of all 4,762 reviews is put in the model as test set. We execute the Naive Bayes method. The result of this method per review consists out of only -1, 0 and 1, where 1 means positive sentiment, 0 means neutral sentiment and -1 means negative sentiment. In total 18.3% of the reviews were assigned a negative sentiment, 12.1% were assigned a neutral sentiment and 69.6% were assigned a positive sentiment.

To validate the results of the judges we perform an intra-rater reliability test. table 10 shows the correlation between the 3 different judges. We observe that there is a lot of correlation between all three judges. Judge 2 has a little less correlation with the other judges, but the correlation is still very high. When there would have been a judge with very low correlation, we should have asked more judges for consistent estimations. For the validation we use all the 952 reviews, including the 19 were the judges were not in agreement.

	Judge 1	Judge 2	Judge 3
Judge 1	1.000	.796	.863
Judge 2	.796	1.000	.759
Judge 3	.863	.759	1.000

Table 10: Intra-rater correlation matrix between the 3 judges

As we mentioned in Section 4.1, the reliability statistic we use is Cronbach's alpha. Cronbach's alpha shows the reliability of the model. It can show values from minus infinity to 1, where only the positive scores have value. The rule of thumb, based on Towns (2013), is that when the Cronbach's alpha is higher than 0.70, the model is reliable. Figure 11 shows an alpha of 0.925 which means our model is very reliable.

Cronbach's Alpha	Cornbach's Alpha based on Standerdized items	N of Items
.925	.926	3

Table 11: Reliability statistic using Cornbach alpha

The intra-class correlation tells us more about the reliability. It describes how strongly units in the same group resemble each other, which we explained in Section 4.1. When observing Table 12 it shows a intraclass correlation of 0.925 which is a very good result. When the intraclass correlation is above 0.7 it is acceptable, if it above 0.8 it is and when it is above 0.9 it is sublime according to Dixon et al. (2001). So we can conclude that our judges for the trainings data are very reliable.

	Intra correlation	Upper bound	Lower bound	Value F test	Sig. F test
Single Meassures	.804	.784	.822	13.320	.000
Average Measures	.925	.916	.933	13.320	.000

Table 12: The intraclass correlation with a confidence interval of 95%

As we mentioned in Section 4.1 we calculated the hitrate calculation and a performance of the model calulation. When we compare the rating and the sentiment appointed by the judges we have a hitrate of 86,7%. This means that the judges and rating agree most of the time.

When the model is executed, it tests the performance on the training set as we mentioned in Section 4.1. The performance of the model is 68.6%, this is the percentage that the model and the judges agree. They disagree most about the neutral sentiment only 19.8%. For the model, this is a difficult group because there are very few reviews in the training set with a neutral sentiment. In general we can conclude that the model performs well. Results of our classification are presented in Table 13. As we can observe, there is definitively more positive sentiment than negative sentiment. The total sentiment per week declines over time and the reviews become more neutral. We will use these results in our panel data model, discussed in the next section.

Week	Neg. Se.	Neg. Se. %	Neutr. Se.	Neutr. Se.%	Pos. Se.	Pos. Se. %	Total UGC
1	226	18.71%	95	7.86%	887	73.43%	1208
2	135	16.88%	75	9.38%	590	73.75%	800
3	89	18.16%	44	8.98%	357	72.88%	490
4	87	19.86%	56	12.79%	295	67.35%	438
5	92	18.18%	73	14.43%	341	67.39%	506
6	80	28.74%	73	17.10%	274	64.17%	427
7	92	21.30%	77	17.82%	263	60.88%	432
8	70	15.18%	83	18.00%	308	66.81%	461
Total	871	18,29%	576	12.10%	3315	69.61%	4762

Table 13: Statistics Sentiment Analysis

Figure 11 shows change in sentiment over time. The graph shows the reviews start with a relatively high fraction of positive and negative reviews. The percentage of negative reviews fluctuates around 20%. The fraction of positive reviews slowly declines over the first 8 weeks, with an exception of week 8. The fraction of neutral reviews starts low and has an inclining trend over the first 8 weeks, with an increase of more than 10% from week 1 to week 8. Overall the reviewers of the games become more neutral, when they write a review longer after the release data of the game.

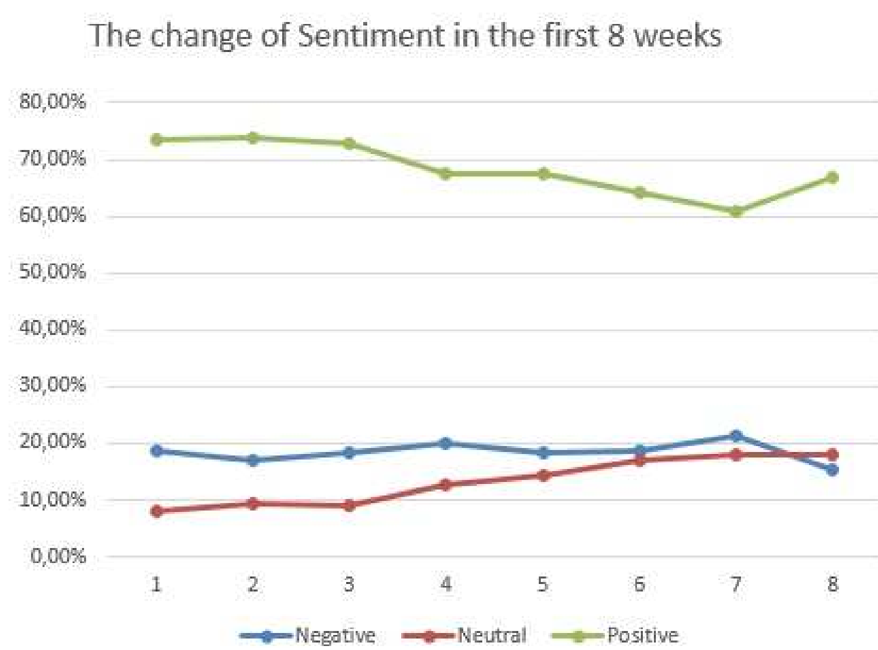


Figure 11: The change of the sentiment over the first 8 weeks after release

For better understanding of our results, Figure 12 is presented. In this figure we observe the proportions of the sentiment of the reviews. We observe a large difference between the positive and the negative reviews in every week of the first 8 weeks after release. As we have seen already, the number of reviews is dropping after more time has passed since the release date. For the number of neutral reviews, there are some fluctations especially in week 3 and 4. Eventually we observe that the number of neutral reviews is increasing, even when the total number of reviews is dropping. The total number of reviews is declining over time as we already discussed in Section 3 and so are the number of reviews with a negative sentiment and positive sentiment, respectively.Reviews, thus, become more neutral over time.

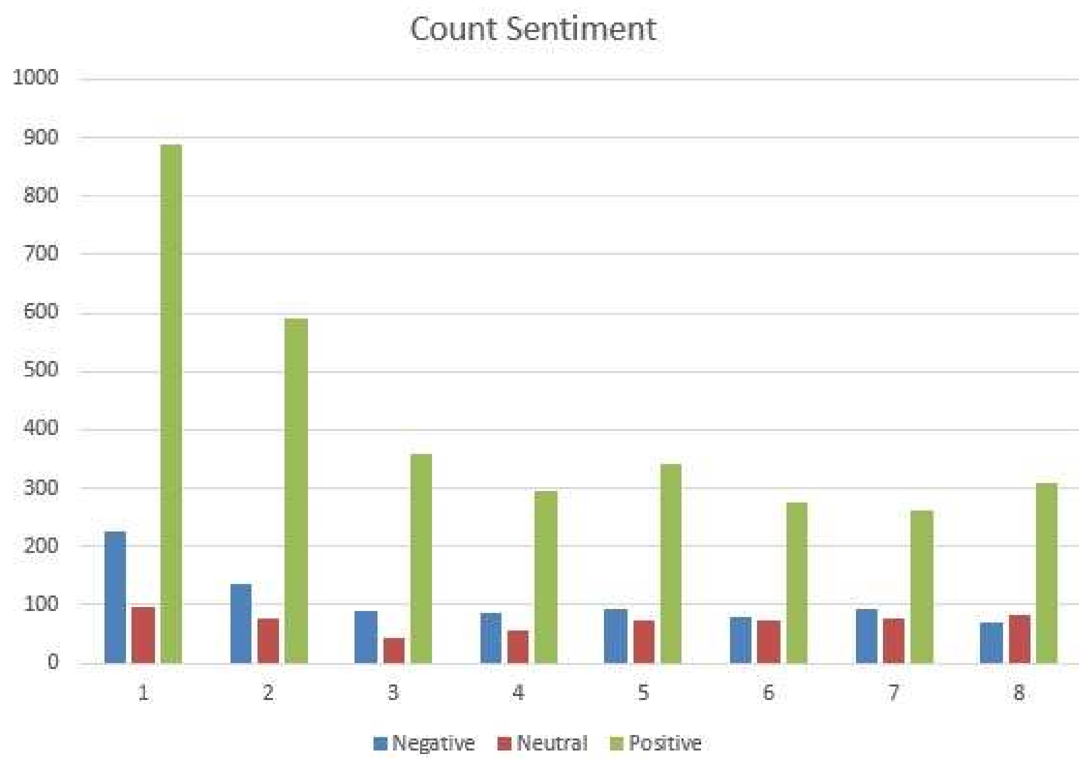


Figure 12: The count of the sentiment over the first 8 weeks after release

When we observe the sentiment over all games, we observe different sentiments for different games. Figure 14 is based on the results of sentiment analysis where we used the Naive Bayes method. We aggregated all reviews per game over all weeks. The percentage negative sentiment of the grand total per game is presented in the second column, the percentage of neutral sentiment is presented in the third column and the percentage of positive sentiment is presentend in the fourth column. Some games have more positive reviews and some games have more negative reviews. The result per game are represented in Figure 14.

Game	Negative (%)	Neutral (%)	Positive (%)	Total (%)
Alien: Isolation	19.23	4.81	75.96	100
Assassin's Creed: Unity	33.11	12.00	54.89	100
Bound By Flame	11.67	1.67	86.67	100
Call of Duty: Advanced Warfare	19.13	15.65	65.22	100
Destiny	23.09	19.19	57.72	100
Dragon Age: Inquisition	14.05	15.70	70.25	100
Far Cry 4	15.89	21.5	62.62	100
Final Fantasy XIV: A Realm Reborn	15.00	1.67	83.33	100
Grand Theft Auto V	17.96	22.54	59.51	100
Guilty Gear Xrd: Sign	16.67	22.22	61.11	100
Just Dance 2015	20.00	15.00	65.00	100
Lara Croft and the Temple of Osiris	28.57	14.29	57.14	100
LEGO The Hobbit	0.00	0.00	100.00	100
LittleBigPlanet 3.00	36.36	9.09	54.55	100
Lords Of The Fallen	10.91	5.45	83.64	100
Metal Gear Solid: Ground Zeroes	22.22	1.19	76.59	100
Middle Earth: Shadow of Mordor	8.99	8.61	82.4	100
Minecraft	7.41	25.93	66.67	100
murdered soul suspect	11.9	9.52	78.57	100
Plants vs Zombies: Garden Warfare	36.36	11.36	52.27	100
Rayman Legends	5.71	0.00	94.29	100
Samurai Warriors 4	9.09	0.00	90.91	100
Sherlock Holmes: Crimes & Punishments	18.18	9.09	72.73	100
Sniper Elite 3	19.35	12.9	67.74	100
The Amazing Spider-Man 2	13.04	4.35	82.61	100
The Evil Within	14.84	5.49	79.67	100
The Last of Us	14.12	16.03	69.85	100
The LEGO Movie Videogame	4.76	0.00	95.24	100
The Walking Dead: Season Two	5.26	15.79	78.95	100
The Wolf Among Us	0.00	20.00	80.00	100
Thief	17.16	0.75	82.09	100
Tomb Raider	6.54	3.74	89.72	100
Total Chatter	0.00	0.00	100.00	100
Warriors Orochi 3	11.11	0.00	88.89	100
Watch Dogs	16.25	7.44	76.31	100
wolfstein the new order	17.54	7.02	75.44	100
WWE 2K15	28.42	11.58	60.00	100
Grand Total	18.29	12.10	69.61	100

Table 14: Sentiment per game

5.2 Correlation and Munticollinearity

All variables are identified at this point. This means we can perform a correlation matrix. The results are presented in Table 15, which shows the correlation between the four variables. When the correlation value is bold that means that the p-value is significant on a 5% significance level.

	Negative	Positive	Volume	Rating
Negative	1.000	**-0.7086**	0.0789	**-0.3640**
Positive	**-0.7086**	1.0000	-0.1080	**0.2235**
Volume	0.0789	-0.1080	1.0000	-0.0747
Rating	**-0.3640**	**0.2235**	-0.0747	1.0000

Table 15: Correlation Matrix independent variables

When we observe Table 15 we can conclude there is correlation between the variables *Rating* and *Negative sentiment*, *Rating* and *Positive sentiment* and *negative sentiment* and *Positive sentiment*. These result confirm our expectation that negative and positive sentiment are correlated, despite the use of neutral sentiment in the sentiment analysis. The reason for this is that they are the opposites of each other and togehter with the neutral sum up to 1. The fact that there is correlation between rating and the sentiment is also logical. The reason is that although rating is a different method of measuring the sentiment than positive and negative sentiment, it still measures sentiment. This leads to correlation with the other sentiment variables.

We use the variance inflation factor (VIF) to determine if there is multicollinearity. There is multicollinearity when it is possible to predict the effect of one variable based on another variable. This phenomenon occurs when there is overlap between variables. This will affect the model because it will correct this overlap, only the effect of the individual variable is much less accurate. Table 16 shows the results of the VIF.

Negative	Positive	Volume	Rating
2.063	2.061	1.0147	1.1593

Table 16: Looking for Multicollinearity using VIF

According to Studenmund (2011) there is multicollinearity when the VIF value is large than 10. When we look at table 16, we can conclude that non of the variables are multicollinear. Thus, this means that we can include all variables in our panel data models.

5.3 Panel data model

As we described in Section 4, we perform panel data models to analyse our data and answer our research question.
By performing our sentiment analysis, discussed in Section 4, we received sentiment for all reviews. By performing this analysis, we received data on sentiment for each game in every week for our dataset. With our panel data models, we will investigate whether this, and our other variables, affects sales. We have performed the Fixed Effects model and the Random Effects model. To decide which model is most consistent with our dataset, we perform a Hausman test.

The Fixed Effects model is defined in the following way in Wooldridge (2010), where $\tilde{X} = X - \overline{X}$ and $\tilde{y} = y - \overline{y}$. This transformation calculates the variables as the deviation of the average of all individual groups.

$$\hat{\beta}_{FE} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{y} \tag{11}$$

The Random Effects model is also based on Wooldridge (2010). For the Random Effects model, it holds that individual effects are assumed to be random. Here the model is rewritten as:

$$y_{it} = \alpha + x_{it}\beta + u_{it} \tag{12}$$

with $u_{it} = \mu_i + v_{it}$ and the individual effects (μ_i) and disturbances v_{it} independent of the independent variables which is in contrast to the Fixed Effects model. $COV(X_{it}, \mu_i) = 0$ and $COV(X_{it}, v_{it}) = 0$ for all i and t.

We performed these models instead of a linear regression model, because of the fact that our data is panel data. Panel data is different from data used in a standard regression model in a way that a datapoint contains both a timestamp and an ID. In our case the timestamp is weekly and the ID represents the game for which the review was written, which we described earlier in Section 3.5. For this reason, we do not only calculate the overall variation, but the between and within variation as well as we described in more detail in Section 4. In Table 17 we observe these three different variations as well as the standard deviation and the mean of all variables, both dependent and independent.

		Mean	VAR	Std
SALES	Overall	9.218	2.759	1.661
	Between		2.072	1.439
	Within		2.560	1.600
VOLUME	Overall	42565.5	544.965	23.344
	Between		371.211	19.267
	Within		402.079	20.052
NEGATIVE	Overall	16.14%	0.0359	0.189
	Between		0.010	0.099
	Within		0.037	0.193
POSITIVE	Overall	73.34%	0.057	0.238
	Between		0.024	0.154
	Within		0.060	0.244
RATING	Overall	4.020	0.553	0.744
	Between		0.318	0.564
	Within		0.675	0.822

Table 17: VAR of the variables

From Table 17 we observe that all variables have a variation above zero. This means that there are no variables for which all values are the same for a specific week (between) and there are games for which all variables have the same value over time (within). From the table, we also observe that for each variable the within variation is higher than the between variation. This implies that there is more variation within a week for a game, than that there is variation between games over all weeks. This could easily be explained by the fact that in the first (two) week(s), a game is highly popular compared to the other weeks. This causes the increase of total sales to decrease over time as we already described in Section 3 as it causes the increase in total volume of UGC to decrease as well. A possible relation with these changes, can also be a shift in fraction of positive or negative reviews per game in more than two weeks after release. Another explanation for the higher within variation for the positive and negative sentiment of UGC over all weeks for our dataset could be the fact that there are more games than weeks, which means that an irregularity of a variable could affect the within variation more than the between variation, because in the within variation, we are simply dealing with less data.

5.3.1 Fixed or Random Effects model

For answering our research question, we need to determine which of the two models is preferred. We performed the Fixed Effects model and the Random Effects model. These models give a slightly different result, because of the different assumptions included in these models. There are different tests to perform that show us the best fit model for our data. We have chosen to use the Hausman test, presented in Hausman (1978) which compares the estimators, $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$, which are consistent under H_0. Under H_1 only the Random Effects estimator is consistent. Thus, under H_0, Random Effects estimator the is preferred, whereas under the alternative the Fixed Effects estimator is preferred. Besides the Hausman test we will use the Mundlak test as well, to verify the result of the Hausman test. In Table 18 we can observe the results of both tests.

Variable	FE	RE	Coef. Diff	S.E. Diff
Negative sentiment	0.687	-0.163	0.850	0.0000
Positive sentiment	0.848	-0.062	0.910	0.0000
Volume	0.031	0.044	-0.013	0.0016
Rating	0.154	0.038	0.117	0.0000
	$\chi^2(4)$	p-value		
Hausman	56.908	0.000		
Mundlak	27.417	0.000		

Table 18: Hausman's test of specification, with resuits of Mundlak

According to Table 18, the Hausman test shows that H_0 is rejected, which means the Fixed Effects model is preferred. According to the Hausman test the Fixed Effects model is the model that is most consistent with our dataset. The p-value shows that the model is valid. To validate the results we obtained from the Hausman test we perform a Mundlak test. The results of the Mundlak test are presented in Table 18 as well.

The Mundlak test agrees with the Hausman test. We should use the Random Effects model when H_0 is accepted. The χ^2-statistic shows a p-value larger than zero, which means we should reject H_0. Thus, the Fixed Effects model is most consistent with our dataset.

As a result we will use the results of the Fixed Effects model instead of the results of the Random Effects model to conclude our research.

5.3.2 Fixed Effects model

As we already discussed, the Fixed Effects model mainly focusses on the within effects, using the following equation:

$$y_{it} = \alpha + x_{it}\beta + \mu_i + v_{it} \tag{13}$$

For this reason, we have performed a between-estimation in addition to the Fixed Effects model as well. In Table 19 we observe the results of the Fixed Effects model. In this table, the variables are represented in the first column, with their coefficient, standard error, t-statistic and p-value. The p-value determines whether a variable has a significant value for the model. We find a variable significant if the p-value is smaller than or equal to 0.05. Below the table the F-statistic is shown, which implies the fitness of the model to our dataset. The table also shows us that we are dealing with an unbalanced panel, which means that the number of datapoint is smaller than the product of the T weeks * N games. For this reason, the model makes its own estimations based on the other games and weeks. We will first discuss the coefficients and their p-value, then we will discuss the model fit by looking at the F-statistic of the model.

Variable	Coeff.	Std. Error	t-stat	p-value
Negative sentiment	0.687	0.449	1.528	0.128
Positive sentiment	0.848	0.377	2.250	0.025
Volume	0.031	0.004	8.062	0.000
Rating	0.154	0.112	1.376	0.170

Table 19: VAR of the variables

Wald F(4, 227) = 18.587, p-value = 0.000, N = 268, n = 37

When we study the coefficient of the Fixed Effects model we observe that the coefficients of all variables are positive. This means that their effect on *Sales* is positive and the increase of one of the variables will cause the sales to increase as well. When we observe the p-value we can conclude that only the variables *Positive sentiment* and *Volume* have a significant value, which means they do influence the sales significantly and are, thus, significantly different from zero. The p-value of *Volume* is even lower than 0.01, which means the Volume is significantly different from zero when using a 0.01% significance level. The F-Statistic has a p-value of 0.000 which is lower than 0.05. This means the model is valid for our dataset.

The Fixed Effects model works in a way that it creates dummies for every ID, or game in our case. μ_i in the above equation represents the individual effect of the games on our data. In Table 20 the individual effects are presented.

ID	μ_i	Std. Error	t-stat	p-value
Alien: Isolation	7.365	0.679	10.843	0.000
Assassin's Creed: Unity	8.197	0.647	12.663	0.000
Bound By Flame	7.209	0.630	11.451	0.000
Call of Duty: Advanced Warefare	9.945	0.631	15.760	0.000
Destiny	7.702	0.674	11.419	0.000
Dragon Age: Inquisition	8.286	0.702	11.809	0.000
Far Cry 4	8.770	0.700	12.522	0.000
Grand Theft Auto V	6.792	0.694	9.787	0.000
Guilty Gear Xrd: Sign	6.449	0.685	9.411	0.000
Just Dance 2015	8.325	0.732	11.371	0.000
LEGO The Hobbit	6.066	0.747	8.126	0.000
Lara Croft and the Temple of Osiris	7.480	0.674	11.093	0.000
LittleBigPlanet 3	8.016	0.765	10.483	0.000
Lords Of The Fallen	7.358	0.668	11.010	0.000
Metal Gear Solid: Ground Zero	7.133	0.661	10.798	0.000
Middle Earth: Shadow of Mordor	7.666	0.732	10.478	0.000
Minecraft	8.499	0.756	11.240	0.000
Plants vs Zombies: Garden Warfare	6.261	0.590	10.612	0.000
Rayman Legends	6.260	0.758	8.255	0.000
Samurai Warriors 4	6.577	0.735	8.952	0.000
Sherlock Holmes: Crimes& punishments	4.728	0.732	6.462	0.000
The Amazing Spider-Man 2	7.045	0.639	11.023	0.000
The Evil Within	8.025	0.653	12.292	0.000
The LEGO Movie Videogame	6.901	0.734	9.403	0.000
The Last of Us	6.668	0.761	8.760	0.000
The Walking Dead: Season Two	6.737	0.696	9.678	0.000
The Wolf Among Us	6.439	0.732	8.795	0.000
Thief	6.785	0.641	10.588	0.000
Tomb Raider	7.016	0.744	9.432	0.000
WWE 2K15	8.451	0.601	14.071	0.000
Warriors Orochi 3	4.520	0.746	6.055	0.000
Watch Dogs	7.802	0.687	11.355	0.000
inFAMOUS: Second Son	8.864	0.726	12.205	0.000
Murdered Soul Suspect	6.305	0.654	9.648	0.000
Sniper Elite 3	6.340	0.670	9.465	0.000
Wolfenstein the new Order	7.272	0.711	10.230	0.000

Table 20: Indivual effect(μ_i) of games(i) in the Fixed Effects model

In Table 20 we observe that the individual effect is significant for all games in our dataset with a p-value of 0.000. This means that all games significantly do have an individual effect on our sales data. The coefficient shows that there are some big differences between games. For example, we observe that the individual effect of the game 'Warriors Orochi 3' is much lower than the individual effect of the game 'Call of Duty: Advanced Warfare'. The difference in this effect is explained by the higher sales for the second game, which is also more popular, but has more reviews as well.

The cause of the individual effect being different between games is mainly due to the ratio $Volume/Sales$. The individual effect is an addition to the constant, and is, thus independent of time. Because the *Volume* is the only independent variable that is not interval scaled, the value can vary a lot between games. On the other hand, *Positive sentiment* and *Negative sentiment* are interval-scaled between 0 and 1 and the *Rating* interval-scaled between 1 and 5. The impact of these variables, together with their coefficients that are smaller than 1, will not impact the total amount on the right handside of the equation as much as the volume does. The ratio, $Volume/Sales$, thus, explains a lot of the individual effect for each game. When we return to our example, the ratio of Call of Duty: Advance Warfare, is, thus, higher than for Warriors Orochi 3. This explanation should hold for all differences in individual effect. Table 21 shows the overall individual effect on the model. It shows that the average individual effect has a coefficient of 7.292, with a standard error of 0.588 which is lower than the individual effect for all separate games. The individual effect, thus, does not vary a lot from each other. The overall individual effect is significant as well, which is self-evident since the individual effects for the separate games are all significant and positive.

id	IndEffect	Std. Error	t-stat	p-value
Overall	7.292	0.588	12.407	0.000 ***

Table 21: Individual effect of games in the Fixed Effects model overall

As we already presented in Section 4.2, the between estimator averages all observations for unit i and then regresses y_i on x_i. This means that in this model, we do not take into account the time unit variation. This is a big difference in contrast to the Fixed Effects model where we look at the within variation. In this way the between estimation is an addition to the Fixed Effects model. The results of the between estimator can be found in Table 22.

Variable	Coefficient	Std. Error	t-stat	p-value
Negative	-8.074	4.422	-1.826	0.077 *
Positive	-4.737	2.537	-1.867	0.071 *
Volume	0.049	0.009	5.155	0.000 ***
Rating	-0.377	0.404	-0.935	0.357
Constant	14.646	3.574	4.098	0.000 ***

Table 22: Between estimator

When we look at Table 22 we are able to conclude that only the constant and the volume have a significant value for the model. The p-value of *Negative sentiment* and *Positive sentiment* are with their 0.077 and 0.071 only significant at 10% significance level, which is not that high. An interesting thing to notice is the fact that the coefficient of *Positive sentiment* on *Sales* would be negative if the variable is significant. This means that the higher the fraction of positive sentiment, the lower the sales, when observing the average of the variables over time. When explaining this fact, we need to use the statistics of the sentiment variables in Section 5.1, where we observed that the fraction of positive reviews declined over time, but the *Volume* did as well. This could imply that after 2 or 3 weeks, the game is not that popular anymore. After weeks of popularity, the total sales keep increasing slowly, but people are more likely to post a review when they are critical on the game, while keep buying the game. This is in contrast to the first weeks, where everyone talks about the hyped game.

For the coefficient of *Volume* in the between estimator, with a value of 0.049, we observe that, again, there is a positive relation between the *Volume* and *Sales*. After comparing the value with the coefficient of the Fixed Effects model, which is 0.031, we can conclude that volume has a larger impact in the between estimator model than in the Fixed Effects model, which could be caused by the other variables not being significant with their high standard errors.

Since our goal is to study in which way UGC is affecting the sales of video games, we have to denote that over time, the *Positive sentiment* and *Volume* do affect *Sales* positively and significantly, while independent of time, only *Volume* affects *Sales* positively and significantly on a 5% significance level.

6 Conclusion

6.1 General conclusion

In this paper we studied how online User-Generated Content (UGC) is related to the sales of video games after the release of such games. We studied this topic with the following research question:

RQ: How does online UGC affect the sales of video games in the first 8 weeks after its release?

To answer this question we had to create a research framework. We used a sentiment analysis to analyse the text of the User-Generated Content, which was in the form of reviews on these games, based on the research of Tirunillai and Tellis (2012). With their research as a framework and the method for creating a test set based on Liu (2006), we chose three judges to classify our training set and perform a Naive Bayes classification to estimate the whole dataset. From this method, we obtained that 69.61 % of the total reviews was positive, in contrast to the 18.29 % of negative reviews in our dataset. The sentiment did not stay the same over the first 8 weeks. In the weeks short after release the consumers were rather extreme, the percentage neutral sentiment in the first week was 7.9%. In the weeks later after the release the sentiment became more neutral, in week 8 the percentage neutral reviews climbed to 18.0%. This meant that positive sentiment as well as the negative sentiment declined in percentage when comparing week 1 to week 8. After the sentiment analysis, we performed a panel data model to analyse the effect of our variables on sales over the first 8 weeks after release, based on the book of Wooldridge (2010).

After the sentiment analysis we have created all our variables that describe the online User-Generated Content. The variables we used in our panel data models are: *Volume*, *Positive Sentiment*, *Negative Sentiment* and *Rating*. When we observed the variance of the variables we could conclude that for all the variables there is a bigger within variation than between variation. This indicates us that there is more difference between the effects on a game in the different weeks of a variable, than there is with other games generally. This is an indication for the use of a Fixed Effects model.

We performed the Fixed Effects model and the Random Effects model and used the Hausman test and Mundlak test to determine which model was preferred for our data. Both tests showed convincingly to use the Fixed Effects model. The results of the Fixed Effects model showed that when we observed the F-statistic, the model was significant. This means the model explained our sales data significantly. The p-value showed that the variables *Positive Sentiment* and *Volume* do have a significant effect on sales. The coefficients of both of the variables is positive, which means that they have a positive effect on *Sales*. The variables *Negative Sentiment* and *Rating* did not have a significant impact on Sales. Besides our independent variables, the Fixed Effects model created the individual effects μ_i for all games as well. Results indicated that all individual effects were significantly different from zero and varied between games from approximately 4.5 until almost 10. This is as we expected, because there are very large differences in the sales numbers as well.

To conclude this paper, first, we divide the online UGC in to four variables. We have seen that *Volume* has a very significant effect on *Sales*. When we look at *Positive Sentiment* we can conclude that is has a significant impact on *Sales* with a significance level of 0.05. The other variables we used to explain the effect of online UGC on *Sales*, where *Rating* and *Negative Sentiment*, not have a significant effect on *Sales* of video games in the first 8 weeks after the release. Thus the online User-Generated content does affect Sales of a video game in the first 8 weeks after release through *Volume* and the *Positive sentiment* of the reviews.

6.2 Academic Contribution and Managerial Implications

Previous to this study, there were no studies that investigated the sales of a game combined with the UGC after the release of a game. There were several papers that found that UGC does have a significant impact on consumer behaviour. For example, Chintagunta et al. (2010) found that online UGC has a significant impact on box office earnings. In Liu (2006), they found that, for movies, expectations before release are high, which may lead to a high rating before release, while in the opening week, people tend to be more critical. Tirunillai and Tellis (2012) found that User-Generated Content has a large impact on the performance of a company on the stock market. They, however, focused on other segments, so with our research we contribute to existing research that investigates the effect of online User-Generated Content in different areas.

With this paper we contribute to existing academic literature in a way that we performed a sentiment analysis on UGC of games and we combined these results with panel data. We showed that online UGC can be analysed with the use of a panel data model. We showed that online UGC has a significant effect on the sales of video games in the first 8 weeks after release. Another contribution to the academic literature is the fact that reviews with a positive sentiment have a significant effect on sales and the reviews with a negative sentiment do not. Besides the reviews with a positive sentiment, we now know that the volume of UGC plays a significant role in the sales of video games. In this way, results and methods do contribute on existing academic literature.

This research could be of great importance for companies as they are able to make better strategic decisions when they use the results provided by this paper. We found that volume of UGC has a significant impact on sales over the first 8 weeks after the release of a game, which they can implement in their strategy by stimulating customers to write reviews on their products. Through this way they could be able to generate more sales, in a very low cost way. Another important discovery is that reviews with a positive sentiment have a significant impact on sales, this means that the game should meet the desires and expectations of the customer from day 1 after the release. Thus, when companies launch their game, customers do not accept it if there are bugs in it. It is not good enough to launch a game on time and update it within two weeks due to the bugs. Another way to accomplish this is to manage the expectations before a game is released. By implementing this, it could be more likely that the consumers will write a positive review on their game.

6.3 Limitations and Directions for Future Research

This research had some very specific limitations. For example, we only looked at games released in the USA, in the year 2014, in the action-genre. For game companies it would be interesting to know if the trends that were shown in the USA were present on other markets as well, like for example the market Europe. We focused only on one genre, because we suspected that (the effect of) UGC could be very different between different genres. Other types of gamers could very well write different types of reviews, because they focus on different things. It would be a great research to reproduce this research with different game genres and, for example, study in which way the effect of UGC differs between different genres.

One of the most important limitations and, thus, a great field for further research is our sales data. The sales data we used for this research was extracted from Vgchartz.com, who create educated guesses of the sales of video games. Actual sales data were not provided online, so an important topic for further research is to use actual sales data from the video game companies, with which it would be possible to assess our variables more accurately which could be used in their strategy better.

A Released Action games 2014

Game Name	# Reviews	Average Rating	Total Sales
Alien: Isolation	104	4.02	127,229
Assassins Creed: Unity	451	3.02	677,521
Bound By Flame	60	3.20	81,478
Call of Duty: Advanced Warfare	230	3.60	1,878,115
Destiny	615	3.45	1,387,712
Dragon Age: Inquisition	242	4.41	432,130
Dynasty Warriors 8: Xtreme Legends	1	5.00	11,236
Far Cry 4	214	4.5	606,161
Final Fantasy XIV: A Realm Reborn	60	4.00	80,211
Grand Theft Auto V	284	4.48	1,709,222
Guilty Gear Xrd: Sign	18	4.5	46,756
Just Dance 2015	20	3.25	85,349
LEGO The Hobbit	10	4.5	27,944
Lara Croft and the Temple of Osiris	7	3	51,890
LittleBigPlanet 3	11	3.82	299,523
Lords Of The Fallen	55	3.73	77,357
Metal Gear Solid: Ground Zeroes	252	3.52	264,404
Middle Earth: Shadow of Mordor	267	4.66	362,017
Minecraft	27	4.85	222,354
Plants vs Zombies: Garden Warfare	44	3.25	30,193
Rayman Legends	35	4.83	31,163
Samurai Warriors 4	11	4.09	28,799
Sherlock Holmes: Crimes & Punishments	11	4.27	4,388
The Amazing Spider-Man 2	46	3.26	67,599
The Evil Within	182	3.58	274,960
The LEGO Movie Videogame	18	4.39	51,430
The Last of Us	524	4.77	441,322
The Lego Movie Videogame	3	5	21,581
The Walking Dead: Season Two	19	4.16	28,891
The Wolf Among Us	15	4.93	22,775
Thief	134	3.28	164,369
Tomb Raider	107	4.71	92,774
WWE 2K15	95	2.85	265,572
Warriors Orochi 3	9	4.56	7,278
Watch Dogs	363	3.94	943695
inFAMOUS: Second Son	31	4.45	627,877
murdered soul suspect	42	3.86	43,617
sniper elite 3	31	3.97	47988
wolfstein the new order	114	4.37	182,403

Table 23: Games with statistics

B Data Extraction

For retreiving our data, first, an Application Programming Interface (API) has to be created, which communicates with the website and obtains the data we want to collect. Because all pages of the reviews on Amazon.com are build in the same way, this API is able to extract data from all similar pages. After creating the API, we only have to collect all URLs from which we want to obtain our data and hit 'Query', after which all URLs are checked and all data is being extracted.

We created two APIs: (1) for the online UGC and (2) the Sales data. A preview of this data-extraction is presented in Figure 13.

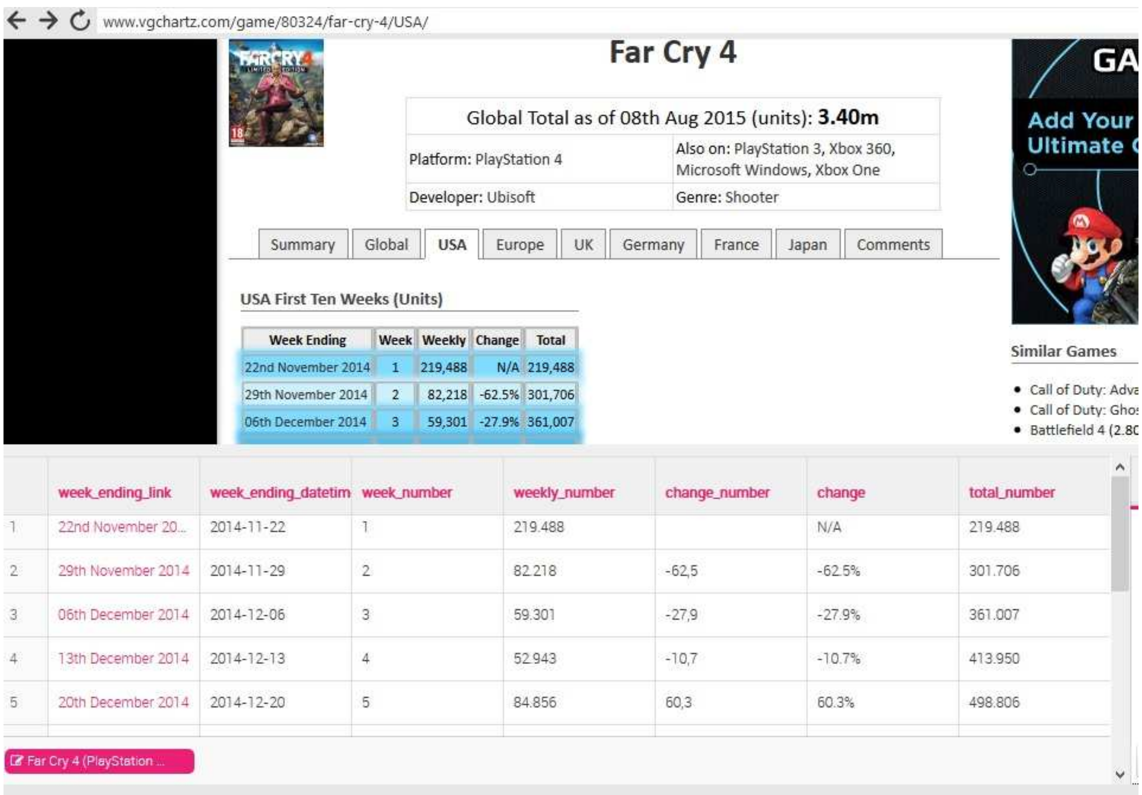


Figure 13: Crawling the web with Import.io

C Process in RapidMiner

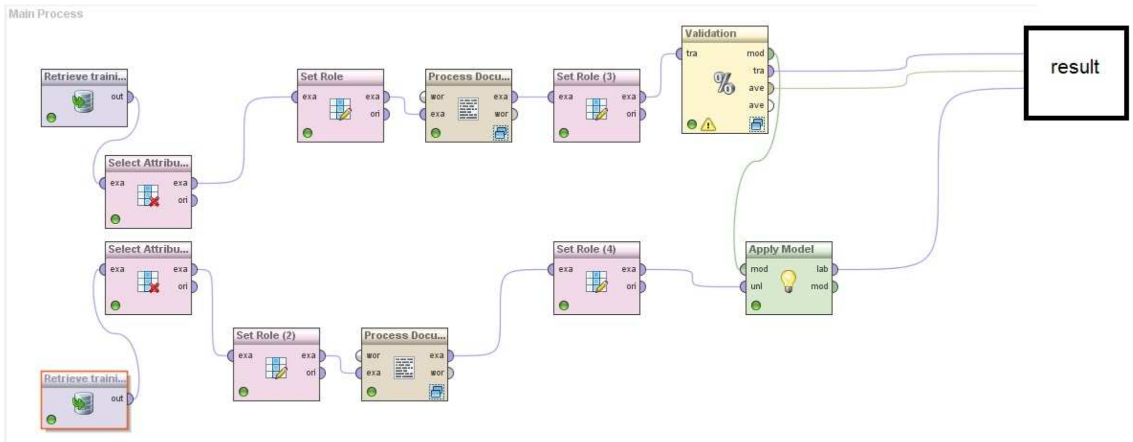


Figure 14: Histogram on the spread of the reviews over all weeks

D Process sentiment analysis

For performing our analysis we will use a software package called RapidMiner. RapidMiner is a software platform that provides an integrated environment for data mining, machine learning and so on. Text mining is also possible with RapidMiner but you have to install a plug-in. We will process the data with RapidMiner so we can use data for a sentiment analysis with the help of RapidMiner.
Figure 14 in Appendix C shows the process of the sentiment analysis in RapidMiner. The process includes the use of the Naive Bayes method.

First of all we need a dataset for training the model. RapidMiner creates the model based on this data. We created our training data with the help of 3 judges, based on the method from Liu (2006). The training set contains 20% of the data of the test set, which were randomly chosen. The 3 judges will give the reviews in the training set a 1 for a positive review, a -1 for a negative review and a 0 for a neutral review. When at least two judges agree the review gets a sentiment as pointed, if no judges are in agreement, the review is removed from the training set. The goal is to create three groups, one positive, one negative and one neutral.The test dataset consists of all 4762 reviews, including the data of the training set. This means that, after the model has been performed, the items that were in the training set, do not necessarily have to have the same sentiment as the judges had assigned. The classification model is, thus, able to overwrite the judges' sentiment for these specific reviews.

The process in RapidMiner, which is discussed more detailed in Appendix C starts with 'Retrieve training set', where the training set is uploaded in RapidMiner. We continue with 'Select Attributes', which selects the relevant attributes, in our case: ID, text and sentiment. We select the attributes and point a role to the ID attribute. After this, we 'process documents to data', which is the preprocessing of the reviews. 'Process documents' contains multiple steps as we can observe in Figure 15.

Figure 15: The process inside Process documents

To process the documents we 'tokenize' the data. The operator splits text in the document into sequences of tokens. The processer can now make a distinction between words, where every word is a token. The next step is 'transform cases', here the processor changes the upper cases to lower cases. We include this because otherwise the processor sees a word that starts with a capital letter or is written in capital letters as a different token than the same word with no capital letters. For the analyzing fase it is important that the processor sees these words as the same. After we tokenized the text and put the text into lower cases, we filter the tokens on text length. The purpose of filtering the tokens on text length is filtering out non-existing words, such as, for example: "amazbalzzzzzzzzzzzzzzzzzzzz" or "Argggggggggggggghhh". The last step in 'process documents' is to filter on stopwords. For example words as: a, is, you, an. These words do not contribute to the sentiment of the text and therefor have to be excluded from the text.

After the sentiment is set as a label in 'set role', we move on to 'validation'. Validation contains different steps as we can see in Figure 16. In the validation the optimal model is developed. We select the attributes we need to optimize the model, here we use the 'text document', we have tokenized and filtered, and the sentiment of the judges.

This sentiment is used to train the model. The tokens will then be transformed from nominal to binominal so we can use the Naive Bayes method. The processor will apply the model several times so the model is able to learn to create an 'optimal? model. To determine which model to use, it uses the performance of the model. Every model that is performed by the processor gets a performance score and the model with the highest performance is the model that best fits our data. This model will be used on the test set in the next step of the process.

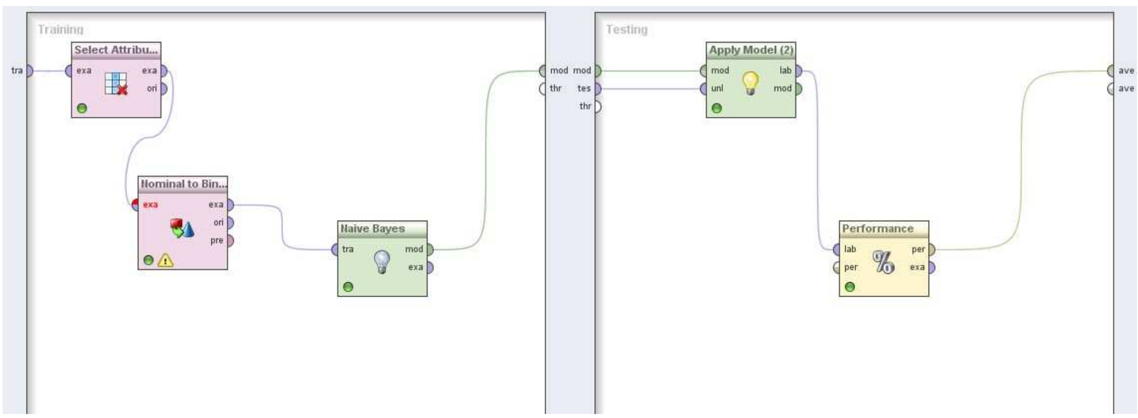


Figure 16: The process inside Validation

The bottom part of the model in Figure 14 in Appendix C works the same way as the upper part of the model. We retrieve the test data in to RapidMiner. We select the attributes we are going to use and we set their roles of the attributes that they are going to have during the process. We process the documents the same way as we did with the training set. We tokenize the reviews, put tokens into lower cases, filter out the tokens based on text length and we remove the stopwords from the tokens. Now that the reviews are tokens we set the roles for the Naive Bayes model we are going to use. The Bayes model we are going to use is selected based on performance on the trainings data in the main process step 'validation'. When we apply the model we retrieve our results. The results will consist of a review with their valence based on the Naive Bayes model.

E Random effects model

The Random Effects model works different from the Fixed Effects model. The underlying assumption implies that (some of) the independent variables will be treated as if they were from random causes. Therefor we are not able to calculate the individual effect. The results of the Random Effects model is summarized in Table 24 .

deptvar	Coefficient	Std. Error	z-stat	p-value
NEGATIVE	-0.163	0.632	-0.258	0.796
POSITIVE	-0.062	0.481	-0.128	0.898
VOLUME	0.044	0.003	12.656	0.000 ***
RATING	0.038	0.117	0.322	0.747
CONSTANT	8.358	0.680	12.292	0.000 ***

Table 24: Results Random Effects Model

Wald $Chi^2(4) = 161.881$, *p-value* $= 0.0000$ When we observe the p-values of the Random Effects model, it shows that only volume and the constant are significant for this model. The other variables rating, negative and positive have very high p-values, which indicates they are significantly not different from zero. There where the p-values in the between model were close to the significant level of 0.05, this is not the case for the Random Effects model. When we look at the coefficients of the variables, we observe that the effect of the positive sentiment is again negative on sales as we with the between estimator. The difference with the between estimator is that in the Random Effects model the coefficients of both the negative and positive sentiment. Both the coefficient in the both models do not have a significant p-value. The coefficient of volume is again positive and is larger than the coefficient in the fixed model.

References

Alvarez Ayuso, I. C., J. Barbero, and J. L. Zofío. A panel data toolbox for matlab. Technical report, Universidad Autónoma de Madrid (Spain), Department of Economic Analysis (Economic Theory and Economic History), 2013.

Archak, N., A. Ghose, and P. G. Ipeirotis. Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8):1485–1509, 2011.

Baltagi, B. *Econometric analysis of panel data*, volume 1. John Wiley & Sons, 2008.

Basuroy, S., S. Chatterjee, and S. A. Ravid. How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of marketing*, 67(4): 103–117, 2003.

Boatwright, P., S. Basuroy, and W. Kamakura. Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics*, 5(4):401–425, 2007.

Bordage, G. and B. Dawson. Experimental study design and grant writing in eight steps and 28 questions. *Medical education*, 37(4):376–385, 2003.

Chen, Y. and J. Xie. Third-party product review and firm marketing strategy. *Marketing Science*, 24(2):218–240, 2005.

Chevalier, J. A. and D. Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.

Chintagunta, P. K., S. Gopinath, and S. Venkataraman. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5):944–957, 2010.

Cui, D. and D. Curry. Prediction in marketing using the support vector machine. *Marketing Science*, 24(4):595–615, 2005.

Das, S. R. and M. Y. Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388, 2007.

Decker, R. and M. Trusov. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4):293–307, 2010.

Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, 49(10):1407–1424, 2003.

Dellarocas, C. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10):1577–1593, 2006.

Dixon, A. L., R. L. Spiro, and M. Jamil. Successful and unsuccessful sales calls: Measuring salesperson attributions and behavioral intentions. *Journal of Marketing*, 65(3):64–78, 2001.

Eliashberg, J. and S. M. Shugan. Film critics: Influencers or predictors? *The Journal of Marketing*, pages 68–78, 1997.

Eliashberg, J., S. K. Hui, and Z. J. Zhang. From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, 53(6):881–893, 2007.

Feldman, R. and J. Sanger. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.

Flanagin, A. J. and M. J. Metzger. Trusting expert-versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, 29(4):1626–1634, 2013.

Franses, P. H. *Time series models for business and economic forecasting*. Cambridge university press, 1998.

Ghose, A., P. G. Ipeirotis, and B. Li. Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3): 493–520, 2012.

Godes, D. and D. Mayzlin. Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4):545–560, 2004.

Griffin, A. and J. R. Hauser. The voice of the customer. *Marketing science*, 12(1):1–27, 1993.

Hausman, J. A. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271, 1978.

Huang, J.-H. and Y.-F. Chen. Herding in online product choice. *Psychology & Marketing*, 23(5):413–428, 2006.

Janssens, W., K. Wijnen, P. De Pelsmacker, and P. Van Kenhove. *Marketing research with SPSS*. Pearson, 2008.

Jun Lee, S. and K. Siau. A review of data mining techniques. *Industrial Management & Data Systems*, 101(1):41–46, 2001.

Klucharev, V., A. Smidts, and G. Fernández. Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. *Social Cognitive and Affective Neuroscience*, 3(4):353–366, 2008.

Lee, T. Y. and E. T. Bradlow. Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5):881–894, 2011.

Liu, Y. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3):74–89, 2006.

Ludwig, S., K. De Ruyter, M. Friedman, E. C. Brüggen, M. Wetzels, and G. Pfann. More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1):87–103, 2013.

Mundlak, Y. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85, 1978.

Netzer, O., R. Feldman, J. Goldenberg, and M. Fresko. Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3):521–543, 2012.

Russell, S. and P. Norvig. Artificial intelligence: a modern approach. 1995.

Shugan, S. M. Editorial: Defining interesting research problems. *Marketing Science*, 22 (1):1–15, 2003.

Singh, S. N., S. Hillmer, and Z. Wang. Efficient methods for sampling responses from large-scale qualitative data. *Marketing Science*, 30(3):532–549, 2011.

Staelin, R. How to write readable papers for marketing science. *Marketing Science*, 25 (3):16–17, 1994.

Studenmund, A. H. Using econometrics: A practical guide. 2011.

Tirunillai, S. and G. J. Tellis. Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science*, 31(2):198–215, 2012.

Tirunillai, S. and G. J. Tellis. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, 2014.

Towns, M. M. Cross-cultural cool: Consumer implications of urban identification in the united states and hong kong. *Journal of Public Policy & Marketing*, 32(special issue): 131–143, 2013.

Wooldridge, J. M. *Econometric analysis of cross section and panel data.* MIT press, 2010.

You, Y., G. G. Vadakkepatt, and A. M. Joshi. A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 79(2):19–39, 2015.

Zhu, F. and X. Zhang. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, 74(2):133–148, 2010.