ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis [Data Science and Marketing Analytics]

"Exploring Gender, Video Game Genres, and Review Helpfulness: A Predictive Modelling Approach"

Name student: Hooijkaas, D. I.
Student ID number: 510268
Supervisor: Bellet, C. S.
Second assessor:

Date final version: 24/07/2023

Abstract
This research explores the intricate relationships between video game genres, gender, and the perceived helpfulness of game reviews using predictive modelling techniques. The aim was to discern whether, and in what ways, perspectives on video games differ by gender. Several models were trained to predict gender from review text and the perceived helpfulness of reviews. Despite challenges in prediction due to the multifaceted nature of these variables, the research provided insightful findings.

The models demonstrated that accurately predicting a reviewer's gender from review text was complex. This complexity challenges traditional beliefs about gender-specific preferences in video games, suggesting a broader and more diverse range of gaming preferences across genders. Additionally, the models employed to predict the perceived helpfulness of reviews highlighted the complex nature of this variable, possibly influenced by factors not included in this study.

Interestingly, the findings provided some evidence for the influence of gender congruence with game genres on the perceived helpfulness of a review, although this influence varied depending on the specific perception of helpfulness and was more genre-specific when considering helpfulness as a count of received votes. This reveals the importance of a nuanced understanding of game genres and their role in predicting gender and perceived review helpfulness.

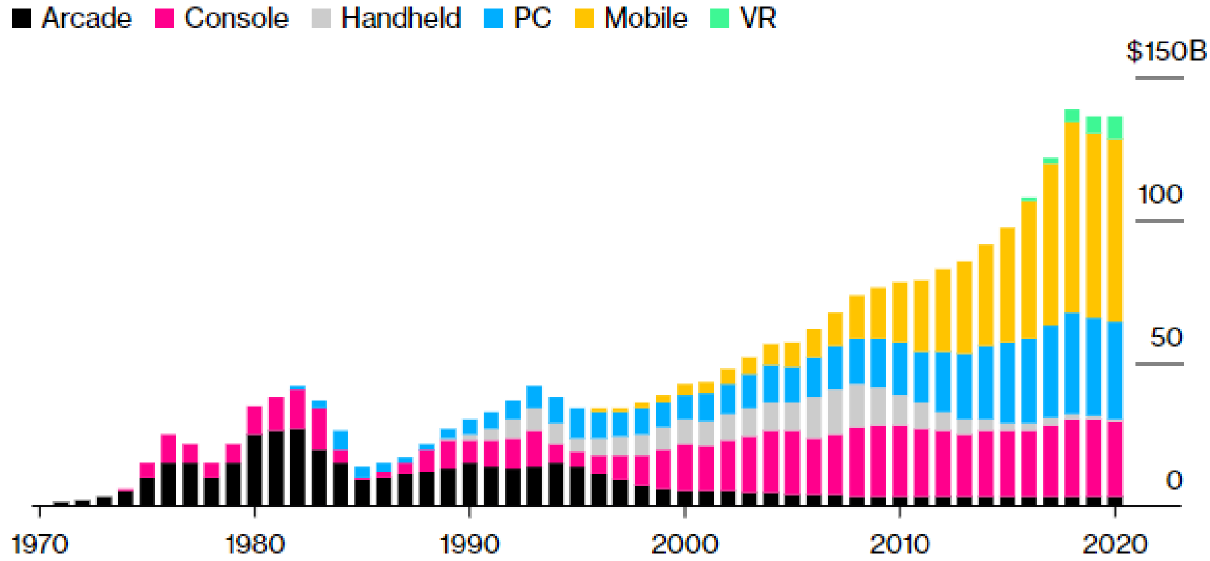
The study recognizes the limitations of the dataset, particularly its gender imbalance, and suggests potential strategies for future research, including the use of broader predictive factors and a more refined perspective on game genres. The findings underscore the complexities inherent in the relationships between gender, game genres, and review helpfulness, thus paving the way for future research that moves beyond traditional gender stereotypes. This nuanced understanding of gaming preferences has important implications for the video game industry, particularly in the areas of marketing and product development.

Contents

Chapter 1. Introduction

The video game market is a relatively new market, having shown rapid growth in the last 2 decades. The video gaming industry particularly saw rapid growth in 2000 with the release of the playstation 2, which still boasts the highest estimated sales of any video game console to this day (Sirani, 2023). Since then, the gaming industry has seen rapid development (The Entertainment Software Organisation, 2022). In modern day, the video game industry appeals to a wide consumer audience, this despite being a more male dominated industry in the past (The Entertainment Software Organisation, 2016). Figure 1 displays the estimated global revenues for the video game industry between 1971 and 2018 (Nakamura, 2019).

Figure 1. Global revenue estimates for the video game industry from 1971 to 2018. Reprinted from "Peak Video Game? Top Analyst Sees Industry Slumping in 2019," by Y. Nakamura, 2019 (https://www.bloomberg.com/news/articles/2019-01-23/peak-video-game-top-analyst-sees-industry-slumping-in-2019#xj4y7vzkg)



Note. Revenues for 2019 and 2020 were future predictions.

Given the rise in popularity of video games, researchers have taken to investigating this market. Research has investigated how game design can be used for marketing purposes (Hamari, & Lehdonvirta, 2010). Understanding the consumer base for video games is crucial for effective marketing. There has been some research on this topic, for instance for games with large quantities of players such as massive multiplayer online (mmo) games (Zackariasson, Wåhlin, & Wilson. 2010). This research will add to this topic by investigating potential gender differences for the consumer audience of video games. Thus this research does not specify itself towards a genre of video games, e.g. mmo games, but rather emphasises a characteristic of the general video game consumer base.

From a video game developer's perspective, information on its consumer audience can be relevant for both game development decisions as well as marketing decisions. As such, it would be interesting to investigate how gender relates to video games. In particular, it would be interesting to see whether gender influences one's view of video games. While some research on this topic exists (Greenberg, Sherry, Lachlan, Lucas, & Holmstrom. 2010), such research is few and far between. Furthermore, given the increasing appeal of video games to larger and more diverse audiences, research regarding the consumer base for video games can quickly become somewhat outdated. As such, this research will add to the existing literature by being both more recent and investigating a niche relation, being that of gender and video games. Ultimately this research aims to answer the question;

"How do the views on video games differ by gender?"

Along with the following sub-questions:
1. How does gender relate to video game preferences?
2. What can be learned from review texts?
3. Do textual product reviews differ between genders?
4. How does gender influence negative reviews?

To accomplish this, this research will analyse review data on video game products from online retailers. First, this research will aim to establish whether there are significant gender differences by training a model to use information from the review and the reviewed product to predict the gender of the reviewer. The idea being that if the model can accurately predict the gender of the reviewer, then there ought to be variables displaying clear differences between genders. From here any resulting differences will be explored. Additionally this research will aim to investigate drivers behind review helpfulness, further exploring potential gender differences with regards to their orientation towards video games. Last, this research will aim to provide relevant advice for video game developers on gender differences in the industry and on the potential for gender targeting.

Chapter 2. Literature

2.1 Video game shopping preferences by gender

Gender has been a relatively major factor influencing consumer behaviour. Gender has been found to influence consumer shopping behaviour, attitudes towards ecommerce, and preferences for in-store shopping (Rodgers & Harris, 2003; Seock & Bailey, 2008). Gender also plays a role in consumers' product preferences. Men and women display differences in shopping preferences for various product categories. One such category is media, which also has a comparable form of consumption to that of video games. In their study on cultural differences, Bertrand and Kamenica (2018) showed men and women displaying different preferences regarding their media consumption. Movies in the categories of action, thriller and sci-fi were predominantly male associated whereas drama and comedy romance were more predominantly female associated. Video games genres can be quite similar to those of movies, where video games genres can also be categorised into for instance action, thriller, romance, or sci-fi. As such it could be expected that the movie genre associations found by

Bertrand and Kamenica would carry over to video games. Indeed there is some evidence that this is the case. A study by Greenberg et al. (2008) looked at orientations towards video games by gender for highschoolers and undergraduates. The results showed that men and women display differing preferences for video game genres. Notably, men show a similar preference for the action genre, specifically racing and sports oriented video games.
Women on the other hand showed a tendency towards puzzles and digital variations of classic board games, being quite different from the romance or drama movie genres. Here the authors argue that the female respondents showed an overarching preference for thoughtful games, whereas competition was a dominant motivator amongst male respondents. Another study into video game preferences by gender found that competition in games negatively impacts females enjoyment of a game (Lucas & Sherry, 2004). It could be argued that these preference findings are in line with gender stereotypes. In a study on public opinion regarding gender stereotypes, Eagly et al. (2020) found that the theme of agency, representing one's own mastery and goal attainment through e.g. ambition and competition, was predominantly associated with men. Communion, representing compassion and expressiveness, was found to be strongly tied to the female gender stereotype with its association to the female gender increasing in more recent years. The male association to agency appeared to be shifting towards a more equal association to both men and women in more recent years. Perhaps the orientations of male and female gamers towards competitive games too will become more similar over time.

Challenge, competition and diversion have shown to be strong motivators for the consumption of video games (Sherry et al., 2012). Challenge and diversion could apply to a variety of video games, both racing games and puzzles could be considered challenging for instance. Competition however is an interesting motivator. While there are some exceptions, competition in video games is generally facilitated through multiplayer, where a multitude of players simultaneously play against each other. As such, games that feature competition are relatively more niche than games that feature challenge or games that act as a diversion. Thus competitiveness is an interesting aspect of video games which has thus far shown differing preferences between men and women. Given that reviews frequently cover the user's opinions of the purchased products, it is likely these preferences will show in the textual product reviews of video game products.

2.2 Information behind textual reviews

To examine gender based differences in online product reviews it is important to first look at what kind of differences could be analysed through review texts. For this, word embeddings can be utilised. The study of word embeddings looks at this by representing semantic relations between words through relationships between vectors in higher dimensions. Using word embeddings it is possible to find word associations to cultural aspects, such as status, affluence, and gender (Kozlowski et al. 2019). Words and their association to specific cultural aspects can also be compared. Thus texts as a whole can convey meaningful relations to such cultural aspects. Furthermore cultural changes over time can also be examined through word embeddings (Garg et al. 2018). Word embeddings thus show to be resilient to shifts over time. As such, reviews can theoretically contain complex information on cultural aspects on a societal level. That said, the general purpose of a review is to convey one's opinion regarding a product or service to potential future customers of said product or service. Reviews are essentially a form of word-of-mouth, where one may endorse or discourage others from purchasing a product or service. Reviews can be quite intricate in their language usage. For instance when a review author endorses a product, readers of that review may find that endorsement more or less persuasive depending on how explicit the endorsement was, as explicit endorsement is perceived by the readers as an indication of expertise from the author (Packard & Berger, 2017). Ultimately textual documents can hold a lot of meaningful information both regarding larger cultural aspects as well as smaller personal aspects regarding the author. Thus textual data such as reviews should prove to hold valuable information.

2.3 Textual reviews by gender

Textual reviews for video game products could differ for many reasons. As mentioned, one such difference could be differing video game genre preferences. Another possibility is that gender influences how video games are viewed. However it is also possible that textual reviews would differ by gender for reasons unrelated to video game products. Men and women have shown to use differing language (Newman et al., 2008). Furthermore, it is likely men and women have different means of describing their attitudes towards products. Women are known to be more expressive than men (Kring & Gordon, 1998). It is thus likely that men and women would show differences with regards to the way they write textual product reviews. A study by Sikdar et al. (2022) looks at how gender signalling in online product reviews relates to the performance of the reviews, as perceived by other users on the platform. In their research, they train various models to predict gender, which was inferred from usernames, using the review text. They were able to obtain roughly 70% to 80% prediction accuracy, implying that there were indeed underlying aspects of the review texts which signalled the authors gender. To relate this back to reviews for video game products, it would thus be likely that video game review texts will differ between genders as well. Sikdah et al. also found that reviews for certain product categories can be perceived as more or less helpful by other users on the platform, depending on the signalled gender of the author. For example, electronics and computer products were generally perceived as more helpful when the review author signalled they were male. This could in part be due to people being less confident in their ideas on topics that fall outside their gender's stereotypical domain (Coffman, 2014). As such, it could be that women are more reluctant to share their ideas for product categories which they perceive to be outside their gender's stereotypical domain. Though it could also be that readers perceive review authors to be more qualified when they write a review for a gender-congruent product category. A study by Forman et al. (2008) examined the role of the reviewer's identity disclosure for product reviews in the electronic markets. In this study, Forman et al. found that reviews which disclosed information about the reviewer's identity through identity-descriptive information received more positive ratings. They suggest that consumers use identity-descriptive information to supplement, or even replace, product information when making purchase decisions and when evaluating the helpfulness of a product review. It could thus be argued that by disclosing the reviewer's gender through gender descriptive information, reviews for certain products may be perceived as more or less helpful. This would explain the findings by Sikdah et al. regarding the relatively higher perceived helpfulness of electronics and computer products by male reviewers. This idea is further supported by a study by Ravula et al. (2013) into the role of gender in the creation and persuasiveness of reviews. In their study, Ravula et al. found that reviews written by women were more authentic but less analytical than reviews by men. They also found that the persuasiveness of a review is influenced by whether the author is perceived to be a male or a female, though this influence differs depending on the product category. Based on these findings, it is possible that reviews for video game products are also perceived as more or less helpful depending on the gender of the author. Sikdar et al. found that reviews for computer products were perceived as more helpful for male reviewers. As such, it could be argued that in the context of video game products as a whole, which are relatively close in nature to computer products, reviews by males are perceived as more helpful. However, given that games can be divided into genres of which some are associated with genders, it would be reasonable to expect that a review's perceived helpfulness as a result of gender congruence will be genre specific. For example, male reviews could be perceived as more helpful for action games, whereas female reviews could be perceived as more helpful for puzzle games.

2.4 The influence of negative textual reviews by gender

Thus far, the main topic has been how product reviews can differ, for instance in the underlying views they convey. However, it is also interesting to consider what kind of impact reviews can have. Or more specifically, do online product reviews influence product sales and if so, what constitutes this influence and could gender play a role? In a meta-analysis covering 26 studies on the relation between sales elasticity and online product reviews, Floyd et al (2014) find evidence that online product reviews do have a significant impact on sales elasticity. Implying that online product reviews can influence sales. Here a leading contributor to a review's influence on sales elasticity is the review's valence, i.e. how positively or negatively a reviewer perceives a product. Here an interesting question is raised, would positive or negative reviews be more influential? In the context of amazon book sales, Chevalier and Mayzlin (2006) find that reviews once again influence sales. However, in their research overwhelmingly negative reviews (1 star) were shown to be more influential on sales than overwhelmingly positive reviews (5 star). Negative reviews being more influential is quite interesting. As discussed in chapter 2.2, men and women tend to use different vocabularies to express themselves. Newman et al. (2008) showed in their study on these different vocabularies that men are more likely to use negations and words that convey negative emotions, especially in speech. Though as previously mentioned, women have shown to be more expressive than men (Kring & Gordon, 1998). Thus it becomes intriguing whether gender has a moderating effect on the relation between review valence and perceived helpfulness. In their study, Craciun et al. (2020) investigated the relation between discrete emotions on review helpfulness with reviewer gender playing a moderating role. Their results indicated that female reviewers are penalised when expressing counter-stereotypical emotions such as anger. Furthermore, female reviewers appear to also be penalised for expressing stereotypical emotions (i.e. anxiety) compared to remaining emotionally neutral. Men on the other hand, are not penalised for expressing stereotypical emotions such as anger, and can in fact benefit from expressing counter-stereotypical emotions such as anxiety. Potentially giving men the upper hand in overwhelmingly negative reviews. A study by Craniun and Moore (2019) on the credibility of negative word of mouth showed that for reviews where there were no reputation cues (e.g. a top reviewer badge), the presence of emotions lowered the credibility of negative reviews by female authors. Interestingly, when reputation cues were present, this relation was flipped and male authors' credibility was lowered when emotions were present. Suggesting male and female reviewers are held to different standards depending on reputation. Given that most reviewers are regular consumers, it could be argued that these findings support the idea that male reviewers have an upper hand when it comes to the perceived helpfulness and credibility of negative reviews.

2.5 Conceptual framework

The literature brought forth some interesting findings. Based on the literature it can be expected that reviews will differ by gender. This is due to gender showing different preferences for video games, different vocabulary, and different forms of expression. As such, it is expected that reviews can be used to accurately predict the reviewer's gender, resulting in the first hypothesis;

H1: Video game review texts can be used to accurately predict the reviewer's gender.

As mentioned, gender is expected to play a role in video game preference. This particularly pertains to the genres which video games can fall under. Here it is expected that genres featuring action and competition will be predominantly associated with male gamers, whereas genres such as puzzles and thinking games will be predominantly associated with female gamers. As such, the following hypotheses can be made;

H2: Video games featuring action and competition will be more associated with men.
H3: Video games featuring thought and puzzles will be more associated with women.

The literature also showed that textual reviews can be perceived differently depending on the reader's awareness of the reviewer's gender. Here it was found that reviews for electronic and computer products, which are arguably product categories commonly associated with the male stereotype, were perceived as more helpful when the reviewer was male. Extending this to video games, it can be expected that reviewers' reviews will be perceived as more helpful for gender congruent game genres. This leads to the fourth hypothesis;

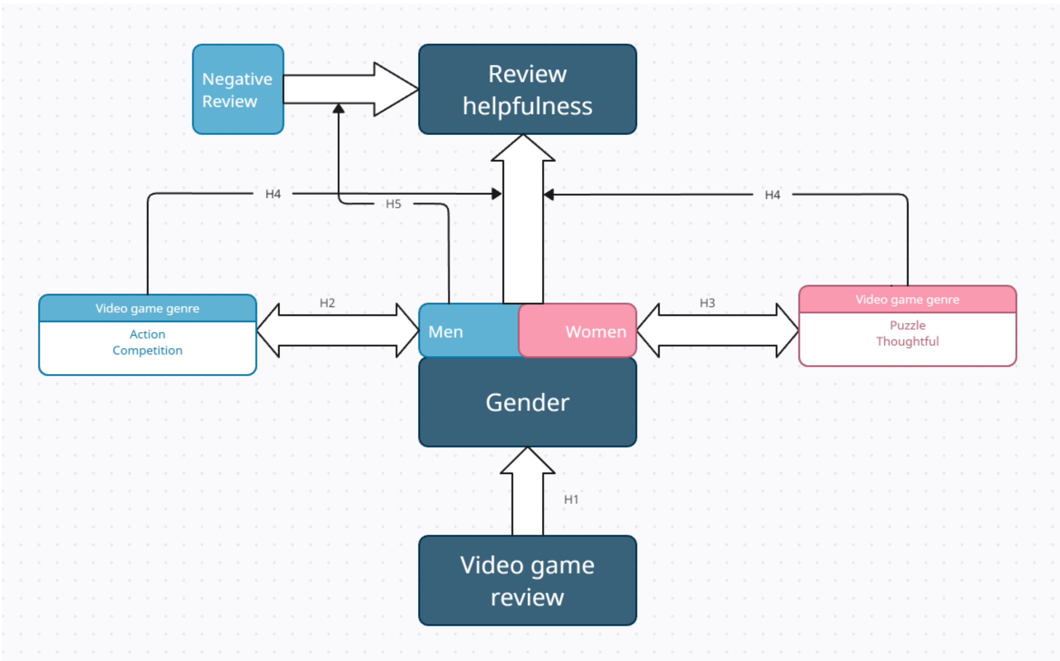
H4: Video game reviews will be perceived as more helpful when the reviewer's gender matches the genre's gender stereotype.

Last, the literature indicated that negative reviews were more influential and that gender played a moderating role in the effectiveness of negative reviews. Here the literature pointed towards men having a leg up with regards to negative reviews, with an exception for high reputation reviewers. Given that this research aims to look at general consumer reviews, it can be expected that negative reviews will be relatively more impactful if the reviewer is male. This leads to the fifth hypothesis;

H5: *Negative video game reviews are perceived as more useful when the reviewer is male*

Ultimately this research's first and foremost interest is to investigate whether video game reviews differ by gender. As such, the predictive power of video game reviews on the reviewer's gender will be investigated. Following this, relations of interest are the relations between gender and video game genres, the relation between gender and review helpfulness, and the relation between gender and the helpfulness of negative reviews. These relations and the hypotheses can be found in Figure 1, which displays the conceptual framework based on the literary findings.

Figure 1. Conceptual Framework



Chapter 3. Data

3.1 Data description

The data for this research is US Amazon review data, featuring reviews between 1996 and 2018. The data has been used in review labelling research (Ni et al. 2019). In particular, for this research the review data on video games will be used. This data on video game products comes in the form of a review dataset, with an additional dataset on amazon products which can be linked back to the review dataset through unique product ID’s. This research focuses on the review dataset, which contains 497577 observations before cleaning. Thus this research does not use the separate dataset featuring the amazon product data.

3.2 Data processing

The processing of review text is a critical step in data analysis, as it prepares the raw text for subsequent analysis by transforming it into a structured and more easily interpretable format. This typically involves several tasks such as general cleaning of interpunction, removing stop words and empty spaces, and applying stemming techniques.

First, it is essential to remove punctuation marks from the review text, such as commas, periods, quotation marks, and other special characters. This is done to ensure that the text analysis algorithms can focus on the words themselves without being influenced by punctuation, which does not typically hold meaningful information in this context. In addition to punctuation removal, it is necessary to remove any extra spaces, including leading and trailing spaces, as well as multiple consecutive spaces. This standardises the spacing within the text, making it more uniform and easier to work with.

Second, it is crucial to remove stopwords from the text. Stopwords are common words such as "and", "is", "in", and "the" which generally do not carry significant meaning and can be safely removed to reduce noise and dimensionality in the dataset.

Stemming is another essential text preprocessing technique that aims to reduce words to their root or base form. By doing so, it enables the analysis algorithms to treat different inflections or derivations of the same word as a single entity, thus improving the accuracy and interpretability of the results. For example, stemming would reduce words like "running", "runner", and "ran" to their root form "run". This process helps reduce the dimensionality of the text data and allows the analysis algorithms to more effectively identify patterns and relationships between words. It should be noted that unstemmed data will be used for sentiment analysis, as stemmed data can lead to a loss of nuance in the context of sentiment extraction.

Furthermore, to infer the gender of the reviewer, first names will be used. For this reason, only reviews featuring a first name can be used. On top of this, any reviews which have become empty after the cleaning of punctuation and the removal of stopwords are removed. Post data cleaning, the dataset features 89731 reviews. Most reviews were lost due to reviewernames not featuring a usable first name due to them either abbreviating their first name, e.g. “J. Smith”, or using a pseudonym, e.g. “ShovelGamer5050”. To check that the exclusion of reviews with unusable names did not introduce bias, a 'sanity check' was performed. This involved comparing a random subset of the excluded reviews with a random subset of the full sample of reviews. T-tests were performed on various characteristics of the reviews, such as ratings, length, emotions, and topics. The results (see Appendix Table B.1) indicated significant differences for emotions, but almost no significant differences in topics or ratings. This introduces a possible limitation to the research as its data may not be a completely accurate reflection of the population. Given that the topics do not show any major selection bias, the resulting data for analysis is still considered adequate for the purposes of this research.

3.3 Exploratory data analysis

Table 1 displays some descriptive statistics on the cleaned video game review data. From Table 1 it can be observed that the gender distribution for the review data is skewed towards males. This should be taken into account when interpreting model prediction accuracy. Table 1 also shows the average rating, which shows females giving slightly higher ratings than males on average. When looking at 1 star reviews, males appear to give ever so slightly more 1 star reviews, though the difference is almost negligible. With 5 star reviews the difference is slightly more apparent, with females appearing to be more inclined to give 5 star reviews than males. Table 1 also shows the average amount of helpful votes given by readers, here it appears that males are given slightly more helpful votes than females on average. Last, Table 1 shows the average review length, measured by the quantity of words used in the review. Here it appears that males use more words than females in their reviews for video game products.

Table 1. Descriptive statistics of the data, by gender.

	Overall	Female	Male
N reviews	89731	22189 (~25%)	67542 (~75%)
Average rating	4.26	4.37	4.22
1 star ratings (% of N reviews)	0.06%	0.05%	0.06%
5 star ratings (% of N reviews	0.63%	0.68%	0.61%
Average helpful votes	1.59	1.56	1.61
Average review length (n words)*	96.75	80.81	101.99

*Note. Values are rounded to two decimals. *word counts were derived from unstemmed reviews, before removing stopwords.*

Figure 2 displays the most frequent words used for the stemmed video game reviews. Unsurprisingly the words “game” and “plai”, are the most frequently used words. Looking beyond these two, the words “time”, “fun”, and “love” appear. Given the relatively high average rating of approximately 4.26 seen in Table 1, it is not too surprising to see positive words such as “fun” and “love” appear frequently. Perhaps “time” is frequent due to people explaining how much time they spent on the video game or mentioning they enjoyed their time playing video games. Some other frequent words are “graphic”, “stori”, and “charact”. These words could indicate that reviews frequently mention or discuss the game’s graphics, story, and characters. These are arguably important aspects of a video game, so these words being frequently used is to be expected.

Figure 2. Word Frequency Histogram

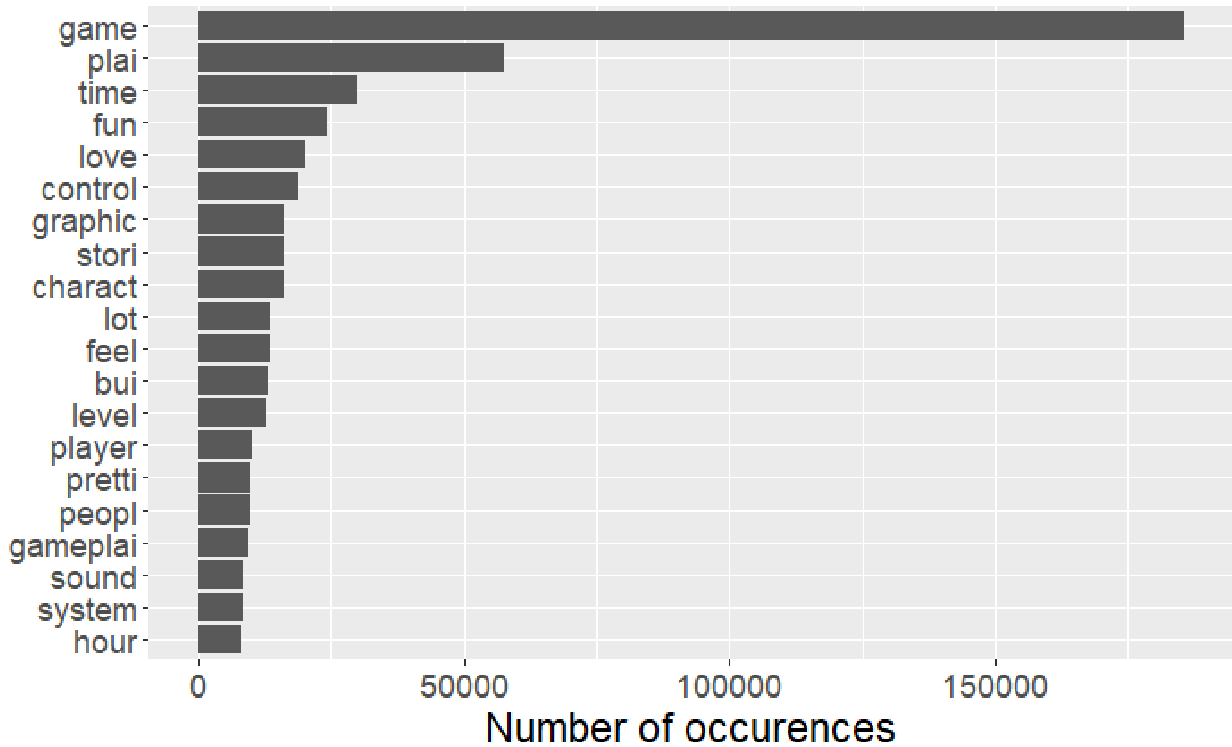


Figure 3 shows a word cloud for frequent word comparison between males and females. From Figure 3 it can be observed that words such as “combat”, “multiplay”, and “fight” are used relatively more frequently by males. These words are more indicative of competitive action games. Females on the other hand appear to use words such as “puzzl”, “sim”, and “family” more. These words are more indicative of puzzle and simulation games. Females also appear to use words such as “gift”, “kid”, and “son”. This could indicate that a portion of the reviews are mothers who bought games for their children.

Figure 3. Comparison Word Cloud between genders.

Chapter 4. Methods

4.1 Analysis objective

This research aims to use the textual data from amazon reviews to investigate whether there is a relation between gender and video games and if so, what constitutes this relation. As such, a quantitative research approach will be taken. To investigate the relation between gender and video games, the gender of the reviewers will have to be inferred as it is not present in the data. Then a machine learning model can be trained on the textual review data to predict the inferred gender. The idea behind this approach is that if the model can accurately predict the inferred gender, then it is likely there is a relation between gender and video game reviews. If this is the case, the model can then also be used to indicate possible drivers.

4.2 Predicting gender

Given that the original data does not feature the reviewer's gender, this will have to be inferred. This will be done via the reviewers' names, using the gender R package (Mullen, 2015). This R package uses historical datasets from various government sources to predict gender based on first names.

4.3 Text transformations

The review texts have to first be processed before they can be properly used to train the machine learning model. Here text can be transformed to word frequencies, word couples frequencies, sentiment emotions, and topics. The frequencies of singular words (unigrams) and word couples (bigrams) is relatively straightforward to obtain by counting their occurrences within a body of text. Sentiments can be acquired using the sentimentr package (Rinker, 2021). This package uses dictionaries to calculate text polarity sentiment. With this it is possible to identify the presence of certain emotions within a body of text, i.e. the word "hate" would increase the likelihood of the emotion of anger being present. Finally, Latent Dirichlet Allocation (LDA) can be used to identify topics (Bei et al., 2003). LDA assumes each document of text within a larger corpus of text is a mixture of topics with words attributed to those topics. Given this, LDA represents topics as probability distributions over the vocabulary of the corpus, meaning each topic assigns a particular probability to every word. Words with the highest probabilities for a given topic are considered the most representative words for that topic. LDA requires the parameter for the number of topics to be set. One approach to finding a good value for this parameter is to run LDA for varying parameter values and compare the interpretability of the resulting topics. Here the parameter which results in the more interpretable topics is preferred. It is possible to use an alternative method such as using a coherence metric (Mimno et al., 2011). Optimising a metric can be done by a machine programme, making it easier in the context of a text corpus featuring a large amount of topics. Additionally it is possible to use word embeddings to extract meaning from words. Word embeddings represent words as dense vectors in a continuous space, capturing semantic relations. Popular models are continuous bag-of-words (CBOW) and skip-gram models in word2vec (Mikolov, Chen, Corrado, & Dean, 2013), and the GloVe (Global Vectors for Word Representation) method (Pennington, Socher, & Manning, 2014). It is possible to train new word embeddings or use pre-trained word embeddings.

4.4 Regression

This research will employ both Poisson and Logistic regressions. Poisson regression is a generalised linear model form of regression analysis used to model count data and contingency tables. It assumes the response variable or outcome follows a Poisson distribution and can be used for modelling the count of events that occur within a fixed volume of time, area, or space. It assumes the log of the outcome variable can be modelled as a linear combination of the predictor variables. The Poisson regression equation is usually expressed as $\log(y) = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp$, where y is the outcome variable, $\beta 0$ is the y-intercept, and $\beta 1$ through βp are the coefficients of the predictor variables X1 through Xp.

Logistic regression, on the other hand, is used when the dependent variable is binary in nature (i.e., two possible outcomes). For example, predicting whether an email is spam or not, or if a tumour is malignant or benign. Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. The logistic function, also called the sigmoid function, can take any real-valued number and map it into a value between 0 and 1, which can be used to represent a probability. The logistic regression equation can be written as $\text{logit}(p) = \log(p/(1-p)) = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp$, where p is the probability of the presence of the characteristic of interest. The coefficients (β values) adjust the input according to their relevance to predict the correct class, mapping the results within the range of 0 and 1.

4.5 Machine learning

Extreme Gradient Boosting (XGBoost) will be used as the machine learning model to analyse the review data. XGBoost, introduced by Chen and Guestrin (2016), is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework. The underlying idea of gradient boosting algorithms is to combine the predictions of several simple models to generate an overall prediction that is more accurate than any of the individual predictions.

XGBoost differs from gradient boosting as it features enhancements and optimizations that make the algorithm faster and more effective. These enhancements include regularisation to prevent overfitting, parallel processing for faster computation, handling of missing values, and the ability to use custom optimization objectives and evaluation criteria.

XGBoost is particularly suitable for large datasets due to its ability to handle sparse data and missing values, as well as its scalability and efficiency. It also provides importance scores for the features, making it useful for feature selection and for understanding which features are most influential in the predictions. As such XGBoost can be applied to predict gender and understand any underlying drivers. Furthermore XGBoost can be applied to predict review helpfulness as well, as decision trees innately incorporate interactions. Thus it is also suitable for analysing how interactions between gender and video game topics and genres relate to review helpfulness. XGBoost can be utilised in R using the XGBoost package (Chen & Guestrin, 2015).

4.6 Limitations

LDA has a few relevant limitations. LDA does not guarantee semantically meaningful or interpretable topics, thus being dependent on the coherence of the words in each topic. As mentioned, LDA also requires a parameter determining the quantity of topics. One approach to this, being relatively straight forward, is to run LDA with varying parameter values to see which topic quantities produce interpretable. This research aims to use LDA topics as input variables for a model. Thus, if a parameter value can be found which produces interpretable topics, then both these limitations are adequately addressed for purposes of this research. Another limitation of LDA is that it assumes a 'bag of words' model, meaning it treats each document as an unordered collection of words. This assumption overlooks the context which can lead to loss of important information. This simplification is part of what makes LDA computationally feasible and conceptually straightforward. To circumvent this assumption, more advanced models like word embeddings and transformer models, designed to capture the context and semantic relationships between words, could be used.
Naturally XGBoost also has some limitations despite its popularity. XGBoost, like other tree-based models, can be biassed in its calculation of feature importance, giving higher importance to variables with more categories or a larger range of values. Though this shouldn’t be too much of an issue in the context of textual analysis as unigrams, bigrams, and topics are treated as separate input variables. XGBoost does use several hyperparameters which need to be tuned. Ideally a huge variety of hyperparameter combinations would be tested to see which perform best, however due to the computational demand this is not always

feasible. As such a smaller set of hyperparameter combinations, appropriate to the available computing power, would be tested.

4.7 Methods summary

This research aims to analyse Amazon video game reviews to investigate a potential relationship between gender and video game preferences. The gender of reviewers is not directly available in the data but will be inferred using their names with the help of an R package. The review text will be processed to obtain word frequencies, sentiment emotions, and topics. An algorithm called Latent Dirichlet Allocation (LDA) will be employed to identify topics within the reviews.

The machine learning model Extreme Gradient Boosting (XGBoost) will be utilised to analyse the processed review data. XGBoost is advantageous due to its speed, efficiency, and ability to handle large, sparse data and missing values. It will be used to predict gender and identify underlying factors driving these predictions. It will also be used to predict review helpfulness, providing insights into how the interaction between gender and video game genres affects review helpfulness. In the case of low prediction accuracy, pre-trained word embeddings could be used in an attempt to improve the model's performance.

The research acknowledges limitations in both LDA and XGBoost. LDA may not always produce meaningful or interpretable topics and overlooks context due to its 'bag of words' assumption. However, for the purpose of this research, these limitations can be mitigated if an appropriate number of interpretable topics can be identified. More advanced models capturing context could also be used. XGBoost might be biassed in feature importance calculations and requires careful hyperparameter tuning. In the context of textual analysis, the bias might be less impactful, and a suitable set of hyperparameters will be tested within the computational constraints.

Chapter 5. Results

5.1 Results overview

In this section the results of this research will be discussed. First the predictors for the model training will be briefly discussed. Then the models trained to predict gender using reviews will be compared. After this the model trained to predict review helpfulness will be analysed and discussed. Last, the results of this analysis will be summarised.

5.2 Predictors

The models are trained using a variety of predictors extracted from the textual review data. The simplest predictors are the number of words in a review as well as unigrams and bigrams, where the 100 most frequent unigrams and the 200 most frequent bigrams were chosen as predictors. Additionally sentiment analysis was used to extract emotion indicators on a review level, resulting in 8 emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). Additionally sentiment analysis was used to create two variables for the overall positivity and negativity of a review. For LDA, as discussed in the methodology, a topic quantity resulting in interpretable topics was chosen by inspecting topics produced by varying topic quantities. Here a topic quantity of 24 was opted for as it resulted in several informative and interpretable topics. Some of these topics include game genres (e.g. horror, strategy, race, story), game features (e.g. multiplayer), game aspects (e.g. sounds, graphics, controls), and even some popular game franchises (e.g. mario, sims, pokemon). The full display of topics with their associated words can be found in the Appendix under Figure A.1.1 These topics were used to obtain topic probabilities on a review level, thus resulting in 24 predictor variables. The topics will be referred to by their interpreted theme, see Appendix Table B.2 for the list of topics and their attributed names. The topics do display differing gender proportions, hinting at possible gender associations to certain topics. This is best displayed by taking the proportions of the genders per topic and then taking the difference between these proportions and the overall gender proportions. Here horror, mission-based action, and character design show an increase in the proportion of females, see Appendix Figure A.1.2. The last predictors were created by aggregating the word vectors obtained through GloVe on a review level. A simple and relatively common method for this is to compute the mean of the word vectors, which is what was done for this research. For GloVe a vector size of 50 was opted for, thus resulting in 50 additional predictor variables. The total quantity of predictors thus being 385.

5.3 Predicting gender

To predict gender, several models were trained and compared to determine the most accurate in predicting the reviewer's gender. Ultimately eight models were trained[1]. The first is a simple logistic regression benchmark model trained on the unigrams, bigrams, sentiment emotions, and LDA topics. The second model is the first XGBoost model trained on the same variables as the logistic regression model. The third model is an XGBoost model and is trained on the GloVe vectors in addition to the variables from the prior models. The fourth model is a small adjustment to the third model, utilising the 'scale_pos_weight' parameter from the xgboost package. This parameter controls the balance of positive and negative weights, which can be useful for unbalanced classes such is the case for gender in the review data. The fifth model is an XGBoost model trained on balanced training data, maintaining a 1:1 ratio of males to females, with the same features as the third model. The remaining three models were trained to investigate change in predictive capabilities when using considerably less variables. The sixth model uses the reviews length, sentiment, and topics. The seventh model uses review length, sentiment, and embeddings. The eighth and final model uses review length, sentiment, topics, and embeddings. Thus the last three models drop the unigrams and bigrams used for the previous models.

These eight models were used to predict on the regular test data and the balanced test data, where male samples were reduced until an equal amount of male and female samples remained. The test data was used as the predictive performance on unseen data is more representative of real-world performance, being the ultimate goal of these models. Table 2 shows the performance metrics of the models' predictions on the unbalanced and balanced

[1] See Appendix Table B.3 for tested hyperparameters

test data sets. It should be noted that gender was transformed into a dummy with 1 representing male and 0 female. From Table 2 it can be observed that the accuracy for the models is not particularly high. This makes sense when looking at the specificity. In the context of these models true negatives represent correctly predicted female reviewers. As such it can be observed that the models specifically struggle to correctly predict the female gender, given the low specificity. This is also reflected in the low AUC-ROC (Area Under the Receiver Operating Characteristic curve) values. This is notably the case for both the predictions on the unbalanced as well as the balanced test data sets. The fifth model, being trained on a balanced training data, appears to be the model most capable of predicting both genders. Given the imbalance in the dataset it is not too surprising to see the models struggle to predict the minority class, being female reviewers in this case. That said, the fifth model's accuracy and balanced accuracy are still on the lower end despite being trained on a balanced training set. The second model appears to have the highest accuracy and F1 score when predicting on the unbalanced test data. The second model almost perfectly predicts the male cases, though it is horrible at predicting the female cases. It also appears that removing the unigrams and bigrams from the models has little impact on the predictive performance. The only notable difference is that model 6 is struggling considerably more to predict female cases than model 2. Though this is not so much the case anymore for models 7 and 8. This could possibly be an indication of the use of unigrams and bigrams leading to overfitting. A potential takeaway from this would be that these results would argue for the importance of quality in predictive variables rather than quantity, given that using fewer variables does not appear to cripple the model. That said, overall none of these models appear to be adequately capable of predicting gender. As such, based on these results, the first hypothesis stating that gender can be accurately predicted using review texts cannot be accepted.

Table 2. Gender prediction models performance metrics

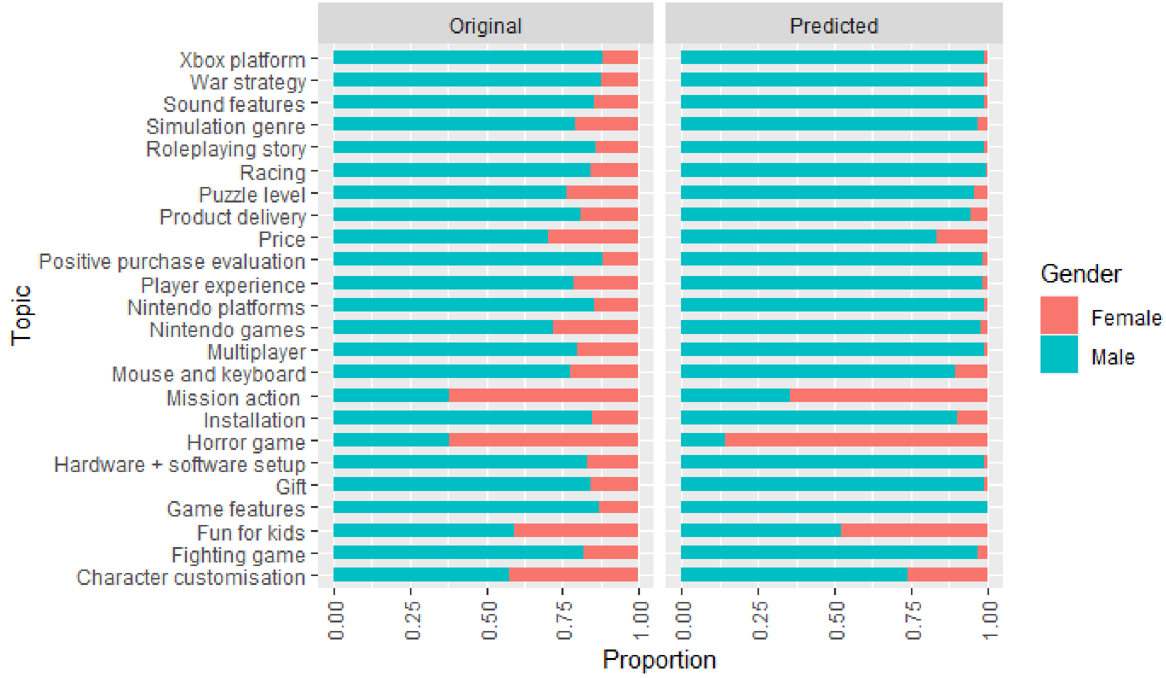
	Model 1 Logistic Regression	Model 2 XGBoost, no embeddings	Model 3 XGBoost, embeddings	Model 4 XGBoost, embeddings, rebalanced weights	Model 5 XGBoost, embeddings, balanced training set	Model 6 XGBoost, sentiment + topics	Model 7 XGBoost, sentiment + embeddings	Model 8 XGBoost, sentiment, topics & embeddings
Unbalanced test data								
Accuracy	0.75	0.80	0.79	0.74	0.68	0.76	0.79	0.79
Sensitivity	0.94	0.99	0.96	0.83	0.71	0.98	0.97	0.96
Specificity	0.18	0.22	0.24	0.47	0.60	0.11	0.24	0.26
Balanced Accuracy	0.56	0.61	0.60	0.65	0.66	0.54	0.61	0.61

F1 Score	0.76	0.89	0.84	0.73	0.67	0.81	0.84	0.84
AUC-ROC	0.44	0.39	0.40	0.35	0.66	0.46	0.39	0.39
Balanced test data								
Accuracy	0.56	0.61	0.61	0.65	0.65	0.54	0.61	0.61
Sensitivity	0.95	0.99	0.97	0.82	0.70	0.97	0.97	0.96
Specificity	0.18	0.22	0.24	0.47	0.60	0.11	0.24	0.26
Balanced Accuracy	0.56	0.61	0.61	0.65	0.65	0.54	0.61	0.61
F1 Score	0.78	0.85	0.82	0.74	0.68	0.75	0.83	0.82
AUC-ROC	0.44	0.39	0.39	0.35	0.65	0.46	0.39	0.39

Note. Values are rounded to two decimals.

So far the overall predictive performance of the models was considered, however this research is particularly interested in the relation between gender and game genres, represented by topics in this analysis. As such, it would be interesting to see how well a model predicts gender per topic. For this particular case, model 8 was opted for. Despite model 2 having slightly better accuracy, model 8 is relatively simpler and has better prediction for female cases. Figure 4 displays the test sample proportions of males and females per topic for the original gender labels and the predicted genders by model 8. From Figure 4 it can be observed that the model underpredicts female cases for male dominated topics. Based on the results from Table 2 it is not surprising to see the model underpredict female cases, it is however interesting to see that it predicts female dominated topics relatively well. Notably the proportion of females for the horror genre increased in the model's predictions. Based on Figure 4, it could be argued that the model predominantly struggles to predict females in the context of male dominated topics.

Figure 4. Original and Predicted Gender Proportions per Topic



Despite the low accuracy, arguably, it can still be interesting to see how these models use variables to make their predictions. Specifically, how the models use certain topics in their predictions of gender can reveal associations the models make between a topic and gender. Based on the literature it was hypothesised that genders would be associated with their stereotypical genres, e.g. males would be associated with competitive action games and females with thought oriented puzzle games. Using partial dependence plots on the relevant topics it is possible to see whether the models do make these associations. Model 2 has the highest accuracy on the unbalanced test data and model 5 has the highest balanced accuracy on the unbalanced test data. Hence the variable associations of these two models are interesting to investigate. For both model 2 and model 5 the most important topics for the model are the topics for the horror genre and the mission-based action genre, see Appendix Figures A.2 and A.5. Based on the literature, it would be expected that mission-based action as a game genre would be more associated with the male gender. However this does not appear to be the case, both models appear to show an association between this topic and the female gender, see Appendix Figures A.6 and A.7.

Similarly the horror genre is also associated with the female gender, see Appendix Figures A.8 and A.9. When it comes to the topic of online multiplayer, the models differ a bit. Model 2 associates online multiplayer with the male gender and model 5 associates a medium probability of the topic presence with the male gender but a high probability of the topic presence with the female gender, see Appendix Figures A.10 and A.11. Interestingly, the models do associate the topics of technical game features, e.g. graphics, and sound features with the male gender, see Appendix Figures A.12 to A.15. This could imply that the models associate more technical topics, which graphics and sound quality arguably are, with the male gender. Another topic associated with the male gender by the models is racing, see Appendix

Figures A.16 and A.17. This topic association is moreso in line with prior expectations based gender stereotypes. Thus far no topics for thought oriented puzzle games have been discussed. This is because the topics from LDA resulted in two thought oriented topics being better described as falling under the strategy genre rather than a puzzle genre. These topics are simulation strategy and war strategy. These topics are interesting as they are both technically strategy games. They mostly differ in that simulation strategy games are more peaceful, where the player often builds and manages cities or civilizations. War strategy games on the other hand generally revolve around conquest and battle simulations, leaning more into the action strategy side. Here the topics do follow gender stereotypes a bit with simulation strategy being associated with females and war strategy being associated with males, see Appendix Figures A.18 to A.21. This could imply that the associations between gender and video game genres is more nuanced than the broader genres such as action or strategy.

Ultimately the results indicate that a genre such as action cannot be solely associated with a single gender. This was indicated by the mission-based action topic being associated with the female gender by the models. Meanwhile racing games, which are arguably a subcategory of action games, were associated with the male gender. Implying that perhaps subgenres are more informative when looking at gender preferences. This point was further supported by the differentiation between topics for simulation strategy war strategy. Here subgenres show differing gender associations from each other. The second hypothesis of this research suggested an association of action and competition genres with males. Conversely, the third hypothesis proposed an association of puzzle and thought genres with females. Based on the results it appears that these associations are more nuanced. While the topics for racing and multiplayer were associated with males, the topic for mission based action games was more associated with females. Hence it would appear that the second hypothesis is only partially supported. Similarly the third hypothesis is also only partially supported as the topic of simulation strategy is associated with females but the topic of war strategy with males. Given the results of this research the second and third hypothesis cannot be accepted. It should be noted that the observed results do indicate that it is possible to make associations between gender and certain genres, though more specific subgenres are likely to yield better results here.

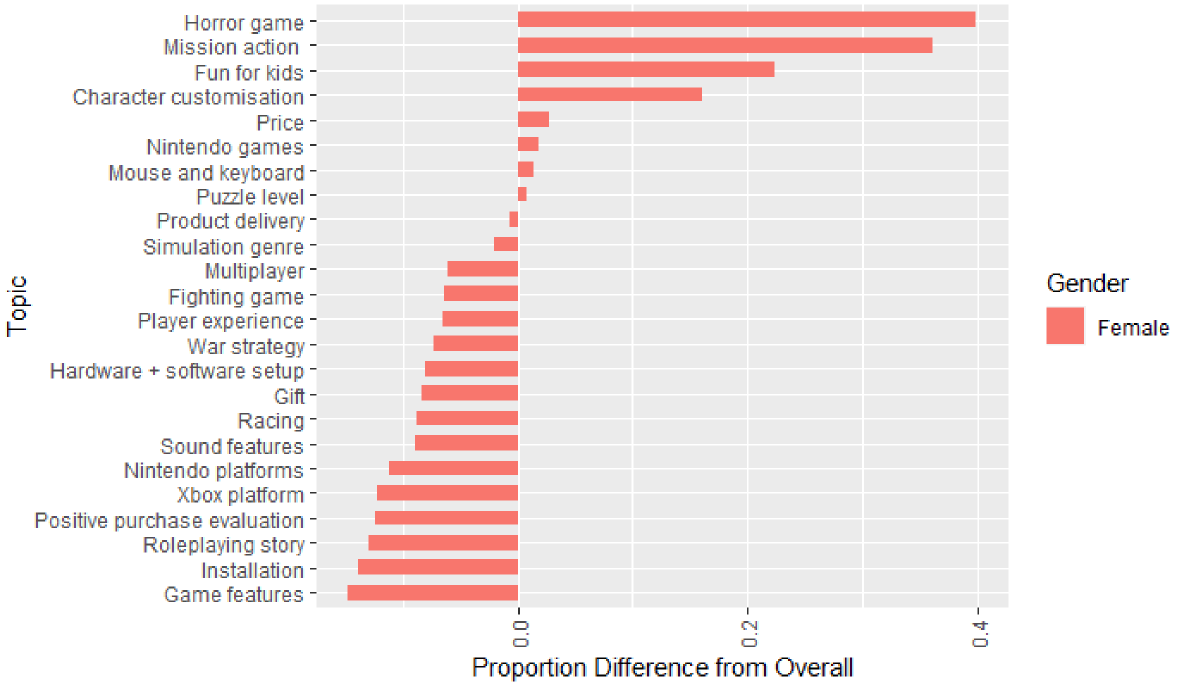
5.4 Review helpfulness

To investigate the relationship between game genres, gender, and review helpfulness several models were trained to predict the helpfulness of a review based on variables extracted from the review text, similar to before. Given the particular interest towards investigating relations between variables, more interpretable regression models were selected for this part of the analysis. The variable chosen to reflect a review's helpfulness is the quantity of (helpful) votes a review received. The distribution of votes for the data sample, being the same data sample from the gender prediction analysis, is quite heavily skewed to the left (see Appendix Figure A.22.). This is the case as roughly 80% of the reviews have received 0 helpful votes. On top of that, there are a few outliers which received incredibly large amounts of helpful votes, being in the thousands. There are only about 200 reviews which received more than

100 helpful votes. Given this, it is not exactly surprising to see the distribution of the votes being so skewed. It is possible to trim the data by excluding reviews which received 0 votes and reviews which received more than 100 votes. The resulting sample would be significantly smaller at 15266 observations (being 85128 observations before). The distribution of the votes is not quite as extreme as previously, but even after trimming the distribution of votes is skewed to the left (see Appendix Figure A.23.). Alternatively, it is possible to transform the variable of interest from a votes count to a helpful dummy variable. The dummy would take value 1 for cases with more than 0 votes, otherwise being 0. This would allow for all the data to be used but would treat all helpful reviews as the same, thus not incorporating how some reviews were perceived as massively more helpful than others. Ultimately several models were trained to cover these approaches.

To investigate the relation between gender and game genres on the perceived helpfulness the regression types Poisson and Logistic were used, incorporating the desired interaction(s) through its independent variables. Given that the quantity of helpful votes received can be considered a count, Poisson was used as the regression for this dependent variable. Naturally Logistic regression was used for the helpful dummy variable. For the independent variables used in the regressions, two approaches were utilised to try and capture gender congruence regarding the video game topics discussed in the review. The simplest approach is to take the dependent variable and regress it on a number of topics, gender, and the interactions between topics and gender. This would investigate the potential influence of gender congruence on a case to case basis. The second approach attempts to aggregate the presence of gender dominant topics in a review, thus ultimately creating measures for the scores of female dominant content and male dominant content within a review. This approach essentially allows one to investigate gender congruence on a review level rather than a topic level. The female and male dominance scores are derived from topic-specific gender proportions, which compare the gender balance within each topic's reviews against the overall gender balance in the entire dataset. For each topic, a positive difference suggests an over-representation of the gender in that topic relative to the overall gender proportion, while a negative difference indicates under-representation (see Appendix Figure A.1.2.). The male dominance score for a review is then calculated by taking the differences between the overall male proportion and the topic specific male proportions for topics where the male gender was over-represented, and multiplying this by the degree to which the topic was present in the review, being the aggregated topic probabilities obtained from lda. The female dominance score was obtained by mirroring the method to apply to the female counterparts. Figure 5 displays the differences between overall female proportions and topic specific female proportions. The topics with the largest positive differences, thus being the most female dominant topics, are displayed at the top with the largest negative differences, thus being least female dominant and most male dominant topics, are displayed at the bottom. The topics for horror games and mission action games being female dominated is not too surprising, the predictions from Figure 4 also showed that the model predicted higher female cases for these topics.

Figure 5. Female dominant topics ranked by gender proportions



Using the female and male dominance scores it becomes possible to incorporate them into a regression model, here the interaction between these scores and gender is particularly of interest. The fourth hypothesis stated that reviews would be perceived as more helpful when the reviewer's gender matches the stereotypically associated gender. Looking at interactions between gender and topic this can be analysed on a case by case level, thus being able to look at each genre separately. Using the interactions between gender and the female and male dominance scores, it is possible to look at gender congruence on a review level. For the models utilising the female and male dominance scores, gender was transformed to have males represented by a "1" and females by a "-1". This was done to allow female cases to have a more explicit influence within the model. Hence the models feature "Male" being a dummy variable for the male gender, and "Gender (recode)" being the transformed gender variable.

Table 3 shows the results for the trained regression models, totalling five models. Models 1 and 2 are Poisson models trained on the full dataset using the votes count as the dependent variable. Model 1 uses the case by case interactions between gender and topics, whereas model 2 uses the interactions between gender and the female and male dominance scores. Model 3 was trained on the trimmed dataset, thus having removed samples where votes were either 0 or larger than 100. Model 3 uses the female and male dominance score interactions. Last are models 4 and 5 which are logistic regression models trained on the full data sample with the helpful dummy, being 1 if votes were non zero (votes > 0) and 0 if not, as the dependent variable. All models feature review rating, review length, and sentiment values as control variables. Table 3 shows the condensed results, the full regression results can be seen in the Appendix Tables A.1 to A.5. For the interpretations of regression results a significance level of 0.05 ($\alpha < 0.05$) will be utilised, represented by a "*" behind the coefficient in Table 3.

From Table 3 it can be observed that most variables used in the models are significant. The first helpfulness model displays the interactions between the gender male and several topics to be of influence in its estimation of the votes count. Genre topics such as war strategy, racing, multiplayer, and mission action appear to be positively related to the male gender. This would imply that a review is estimated to receive more helpful votes if the reviewer is male for video game reviews covering these topics. The two most female dominated topics from Figure 5 were horror games and mission action games, so it is quite interesting to see mission action being positively related to the male gender. The topic of horror games on the other hand displays a strong negative relation to the male gender. This implies that reviews for horror games are estimated to receive considerably more helpful votes if the reviewer is female. Two other surprises were the interactions between simulation and the male gender, and between the sound topic and the male gender. Based on the associations between gender and topics made by the gender prediction models, it would be expected that the male gender shows a negative relation to the simulation genre and a positive relation to the sound quality topic, yet the opposite is the case for the first helpfulness model. For the simulation genre it could be that despite women playing the genre, people still stereotypically associate it with the male gender. This could be one explanation for the males receiving more helpful votes on this genre while the gender prediction models did pick up the genre's association to the female gender. Figure 5 shows the simulation genre topic to be around the middle ground between female and male dominant topics, giving some credit to the idea, though naturally these results could be due to another reason. For the sound quality topic this line of thinking does not seem as credible as Figure 5 showed the topic to be quite male dominated.

Looking at the second model, which uses the male and female dominance scores, it can be observed that the interactions between gender and the dominance scores deviates from prior expectations. The interaction between gender and the female dominance score does not appear to be significant for this model. This would imply that a review covering female dominated topics is not estimated to receive more or less helpful votes depending on the reviewer's gender. The interaction between gender and the male dominance score does appear significant though the coefficient is negative. This would imply that a review covering male dominated topics is estimated to receive more helpful votes if the reviewer is female instead of male. This contradicts prior expectations.

The third model features the same variables as model 2, but was trained on the trimmed dataset. The third model shows the opposite interaction effect for the interaction between gender and male content dominance with it being positive this time. Notably the interaction between gender and female content dominance is significant this time around, also showing a positive interaction effect. Hence it is implied that when looking at reviews for which the received votes count lies on the interval (0,100], both reviews covering female dominant topics as well as male dominated topics are estimated to receive more helpful votes if the reviewer is male. This is a somewhat surprising change from the second helpfulness model, though it can likely be explained by roughly 80% of the observations being dropped in the trimming. Helpfulness models 4 and 5 utilise the transformed helpful variable, being a dummy variable for reviews with more than 0 votes received. This somewhat combats the skewed votes distribution, though the data is still imbalanced as only roughly 20% of the reviews had more than 0 votes received. This is reflected in the somewhat adequate accuracy, but poor balanced accuracy for both models.

Looking at the fourth model it is quite notable how many of the previously significant interactions between the male gender and the genre topics are no longer significant for model 4. Only the genre topics for horror and war strategy are significant this time around, with both displaying negative coefficients. This would imply that for both these topics the review is estimated to receive more helpful votes if the reviewer is female. For the horror genre this is expected, but seeing war strategy be negatively related to the male gender here is a bit surprising. Due to the different dependent variables, naturally there is a difference in what the models capture. The fourth model merely captures whether a review received any amount of helpful votes, whereas the first model also distinguished between the amounts of helpful votes received. As such it could be that an interaction between a game genre and gender has little to no influence on whether a review receives that first helpful vote, but does influence how many votes are received once that threshold of at least one vote has been broken. This would explain why some genre gender interactions are significant in the first model, but not the fourth. The same line of thought could be used to try and explain the sign change for the coefficient of the gender interaction with the war strategy genre, though admittedly this may not be the best explanation for this rather surprising observation.

Finally, the fifth helpfulness model shows the interactions between gender and both the male as well as the female content dominance scores to be significant. Here the interaction between gender and male content dominance sees a positive coefficient and the interaction between gender and female content dominance sees a negative coefficient. This would imply that the fifth model estimates a review to be more likely to receive any amount of helpful votes larger than 0 if the reviewer's gender matches the dominant content gender. Implying that gender congruence would be influential in the review being perceived as helpful, but not really in the review receiving a large amount of helpful votes. Specific interactions between gender and certain game genres appear more influential in a review receiving a large amount of votes than overall gender congruence.

Based on these results there is some adequate support in favour of the fourth hypothesis. Despite some results contradicting expectations and implying the need for a more nuanced look, notable evidence has been shown to support the hypothesis that reviews are perceived as more helpful when the reviewer's gender matches the stereotypically associated gender for a game's genre. Particularly the interaction between horror games and gender would insinuate that females are perceived as quite drastically more helpful when reviewing horror games, given the relatively strong relation displayed. Unfortunately the models do show some struggle in their predictions. The poisson models appear to struggle in predicting cases with higher vote quantities (see Appendix figures A.24 to A.26). The logistic models show an adequate prediction accuracy around 80%, though the balanced accuracy drops to around 60%. Thus it could ultimately be argued that the resulting relations brought forth by the models are somewhat unreliable as a result of these prediction inadequacies. That said, given the relative consistency of the relations across the models, it could be argued that there is enough evidence to support that these relations play a role in the prediction of votes while admitting that the impact and perhaps the significance of these relations may shift slightly when investigated through alternative models.Thus, it was ultimately decided that there was enough evidence to support accepting the fourth hypothesis. Gender congruence did appear to be influential to some degree in the perceived helpfulness of a review, though it would appear that the mediating role of gender differs depending on the context. In the context of receiving any amount of helpful votes, gender congruence appears to have a broader impact, influencing gender dominant content as a whole. In the context of one review receiving more votes than another, gender congruence appears more influential on a game topic specific level. This distinction should be noted.

Last there is the fifth hypothesis on the interaction effect of the male gender and negativity on the perceived helpfulness of a review. Both the first and the fourth model show the interaction effect for the male gender and the review rating to be significant and negative, implying that a review with a higher rating is perceived as less helpful if the reviewer is male. For the interaction effect between the male gender and negative sentiment, the coefficient appears significant and negative. This implies that a review with a negative sentiment is perceived as less helpful if the reviewer is male. Given that both models indicate the same interaction results, the fifth hypothesis cannot be supported on the grounds of these results.

Table 3. Linear regression models predicting review helpfulness votes (shortened results)

	Helpfulness model 1 Poisson regression, case by case gender congruence, full data	Helpfulness model 2 Poisson regression, overall gender congruence, full data	Helpfulness model 3 Poisson regression, overall gender congruence, trimmed data	Helpfulness model 4 Logistic regression, case by case gender congruence, full data	Helpfulness model 5 Logistic regression, overall gender congruence, full data
	RMSE: 13.4	RMSE: 13.7	RMSE: 10.2	Accuracy: 0.84 Balanced accuracy: 0.61	Accuracy: 0.84 Balanced accuracy: 0.60
	β coefficients	β coefficients	β coefficients	β coefficients	β coefficients
Rating	-0.18***	-0.17***	-0.06***	-0.25***	-0.27***
Male	-0.05			0.70***	
Male * rating	-0.01*			-0.04*	
Gender (recode)		-0.06***	-0.09***		-0.10*
Gender (recode) * Male content dominance score		-0.07**	0.10**		0.21*
Gender (recode) * Female content		0.03	0.14***		-0.38**

dominance score					
Negative	-0.01***	-0.01***	-0.01***	0.08***	0.00
Male * negative	-0.01***			-0.10***	
Anger	-0.02***	-0.02***	-0.01**	-0.00	-0.00
Sadness	0.04***	0.04***	-0.01***	0.01	0.01
Positive	0.07***	0.07***	0.03***	0.03***	0.03***
Joy	-0.05***	-0.05***	-0.01***	0.01	0.00
Anticipation	0.05***	0.05***	0.02***	0.08***	0.08***
Horror topic * Male	-19.75***			-21.90***	
Mission action topic * Male	0.55***			0.17	
Simulation topic * Male	0.45***			0.15	
War strategy topic * Male	1.16***			-0.66**	
Racing topic * Male	0.31***			-0.09	
Multiplayer topic * Male	0.47***			0.38.	
Graphics topic * Male	0.38***			0.80*	
Sound topic * Male	-0.49***			0.03	

*Note. Gender (recode) takes value 1 for males and -1 for females. Values are rounded to two decimals. P-values * < 0.05, ** < 0.01, *** < 0.001.*

5.5 Results summary

In this results chapter, the two primary areas of focus were gender prediction and review helpfulness prediction.

The research began with gender prediction, employing both topic and language variables to predict a reviewer's gender. Several XGBoost models were used, which highlighted the significance of topic variables over language variables. Game genre topics, such as 'mission-based action games' and 'survival horror games,' displayed the most considerable predictive value. The emotional tone of the review was a significant language variable. While the models achieved a decent accuracy, they had limitations due to the imbalance in the dataset, with a majority of the reviews written by males. This led to a skewed prediction favouring the male gender. Using a balanced dataset resulted in a model that was better equipped to predict female reviews but showed reduced overall accuracy.

The analysis then moved to review helpfulness prediction, using variables such as the reviewer's inferred gender, rating given by the reviewer, unigrams, bigrams, emotions, and LDA topics. The model's performance was not exceptional, particularly struggling with predicting reviews receiving larger vote counts. Nevertheless, it uncovered potential correlations between variables and predicted received votes.

Here two hypotheses were tested: that reviews are considered more helpful when the reviewer's genre aligns with the genre's gender stereotype, and that negative reviews would be deemed more helpful if written by a male reviewer. The results provided an adequate amount of evidence to support the hypothesis that gender congruence, with regards to game genres, influences the perceived helpfulness of a review. Though it should be noted that this influence varies depending on the defined perceived helpfulness, playing a broader role when looking at perceived helpfulness as a dummy and playing a more genre specific role when looking at helpfulness as a count of received votes. Last, the data however did not provide substantial evidence to support the hypothesis that negative reviews are perceived as more helpful for male reviewers.

Chapter 6. Conclusion and Discussion

This research set out to investigate the relationships among game genres, gender, and the perceived helpfulness of game reviews using predictive modelling. The goal was to answer the question: “How do the views on video games differ by gender?”. Several models were developed to predict gender and review helpfulness, to explore potential systematic differences and the role of gender in these contexts. Although the models had limitations, they provided valuable insights into the examined relationships.

The research findings showed that predicting a reviewer's gender based on their review text was challenging due to a multitude of factors. This complexity indicates a shift away from traditional beliefs about gender-specific preferences in video games, suggesting a growing diversity in gaming preferences across genders. The results revealed the predictive value of certain game genres over others in the context of gender, with specific topics like 'mission-based action games' and 'survival horror games' demonstrating a strong predictive ability.

In predicting review helpfulness, some difficulties were encountered, particularly with regards to the skewed distribution for the variable representing the received quantity of helpful votes, leading to some inaccurate predictions for the tail end of the skewed data. However, correlations between variables and predicted received votes were uncovered, suggesting the presence of complex relationships among these variables.

Interestingly, the research findings provided some support for the hypothesis that gender congruence with regards to game genres influences the perceived helpfulness of a review, though further research to solidify these findings is encouraged. However, this influence varied depending on the defined perceived helpfulness and was more genre-specific when looking at helpfulness as a count of received votes. This demonstrates the need for a nuanced understanding of game genres and their role in gender prediction and perceived review helpfulness.

Future research should delve deeper into this, possibly developing a taxonomy of sub-genres and investigating the distribution of gender preferences across these detailed classifications. This refined perspective could be a valuable tool for the video game industry in tailoring and marketing their products. Future research might consider a broader range of predictors such as user profiles, gameplay hours, or user review history to enhance prediction models.

The limitations of this study, including the gender imbalance in the dataset and potential selection bias, should be acknowledged. Possible alternative solutions to these issues include gathering information through surveys or including demographic factors beyond gender. It should also be noted that the study's reliance on English language reviews might have introduced cultural bias, limiting the generalizability of the results.

Overall, this study highlights the complexity of the relationships between gender, game genres, and review helpfulness. It provides valuable insights for companies developing and marketing video games, particularly highlighting the importance of nuanced genre categorization in considering marketing segmentation. When it comes to sponsored reviews, gender congruence may play a role depending on the genre of the video game, providing some direction for marketing efforts. This research opens avenues for further research into the areas of gender and video games and calls for a nuanced understanding of gaming preferences, moving beyond traditional gender stereotypes. The inclusion of more detailed characteristic features could be informative in trying to analyse and understand consumer tendencies in the video game market.

Bibliography:

Bertrand, M., & Kamenica, E. (2018). Coming apart? Cultural distances in the United States over time (No. w24771). *National Bureau of Economic Research.*

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3,* 993-1022.

Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: Extreme gradient boosting. R package version 0.4-2. https://cran.r-project.org/web/packages/xgboost/index.html

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).* https://doi.org/10.1145/2939672.2939785

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

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics, 129(4),* 1625-1660.

Craciun, G., & Moore, K. (2019). Credibility of negative online product reviews: Reviewer gender, reputation and emotion effects. *Computers in Human Behavior, 97,* 104-115.

Craciun, G., Zhou, W., & Shan, Z. (2020). Discrete emotions effects on electronic word-of-mouth helpfulness: The moderating role of reviewer gender and contextual emotional tone. *Decision Support Systems, 130,* 113226.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American psychologist, 75(3),* 301.

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of retailing*, 90(2), 217-232.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research, 19(3),* 291-313.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115(16), E3635-E3644.*

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming, 41(2), 238-259.*

Hamari, J., & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science & Applied Management, 5(1), 14-29.*

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review, 84(5), 905-949.*

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology, 74(3),* 686.

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication research, 31(5),* 499-523.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of the International Conference on Learning Representations (ICLR 2013), Workshop Track.*

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP '11), (pp. 262-272). Association for Computational Linguistics. https://aclanthology.org/D11-1024/

Moon, S., Park, Y., & Seog Kim, Y. (2014). The impact of text product reviews on sales. European Journal of Marketing, 48(11/12), 2176-2197.

Mullen, L. (2015). gender: Predict Gender from Names Using Historical Data. R package version 0.5.0.9000. https://github.com/ropensci/gender

Nakamura, Y. (2019, January 23). *Peak Video Game? Top Analyst Sees Industry Slumping in 2019.* Bloomberg. https://www.bloomberg.com/news/articles/2019-01-23/peak-video-game-top-analyst-sees-industry-slumping-in-2019#xj4y7vzkg

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse processes, 45(3),* 211-236.

Ni, J., Li, J., & McAuley, J. (2019, November). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. *In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) (pp. 188-197).*

Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research, 54(4),* 572-588.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), (pp. 1532-1543). Association for Computational Linguistics.*

Ravula, P., Bhatnagar, A., & Gauri, D. K. (2023). Role of gender in the creation and persuasiveness of online reviews. *Journal of Business Research, 154, 113386*.

Rinker, T. W. (2021). sentimentr: Calculate Text Polarity Sentiment version 2.9.0. https://github.com/trinker/sentimentr

Rodgers, S., & Harris, M. A. (2003). Gender and e-commerce: An exploratory study. *Journal of advertising research, 43(3),* 322-329.

Seock, Y. K., & Bailey, L. R. (2008). The influence of college students' shopping orientations and gender differences on online information searches and purchase behaviours. *International journal of consumer studies, 32(2), 113-121.*

Sherry, J. L., Greenberg, B. S., Lucas, K., & Lachlan, K. (2012). Video game uses and gratifications as predictors of use and game preference. *In Playing video games (pp. 248-262).* Routledge.

Sikdar, S., Sachdeva, R., Wachs, J., Lemmerich, F., & Strohmaier, M. (2022). The effects of gender signals and performance in online product reviews. *Frontiers in big Data, 4, 118.*

Sirani, J. (2023, Month Day). Best-selling video game consoles: Nintendo, PlayStation, PS5, Xbox. IGN. https://www.ign.com/articles/best-selling-video-game-consoles-

nintendo-playstation-ps5-xbox
The Entertainment Software Organisation. (2016). *2016 Essential facts about the video game industry [fact sheet].* https://library.princeton.edu/sites/default/files/2006.pdf
The Entertainment Software Organisation. (2022). *2022 Essential facts about the video game industry [fact sheet].* https://www.theesa.com/wp-content/uploads/2022/06/2022-Essential-Facts-About-the-Video-Game-Industry.pdf
Zackariasson, P., Wåhlin, N., & Wilson, T. L. (2010). Virtual identities and market segmentation in marketing in and through massively multiplayer online games (MMOGs). *Services Marketing Quarterly, 31(3), 275-295.*
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing, 74(2), 133-148.*

Appendix A - results figures and tables

Figure A.1.1 LDA topics

Figure A.1.2. Difference of Gender Proportions per Topic from Overall Proportions

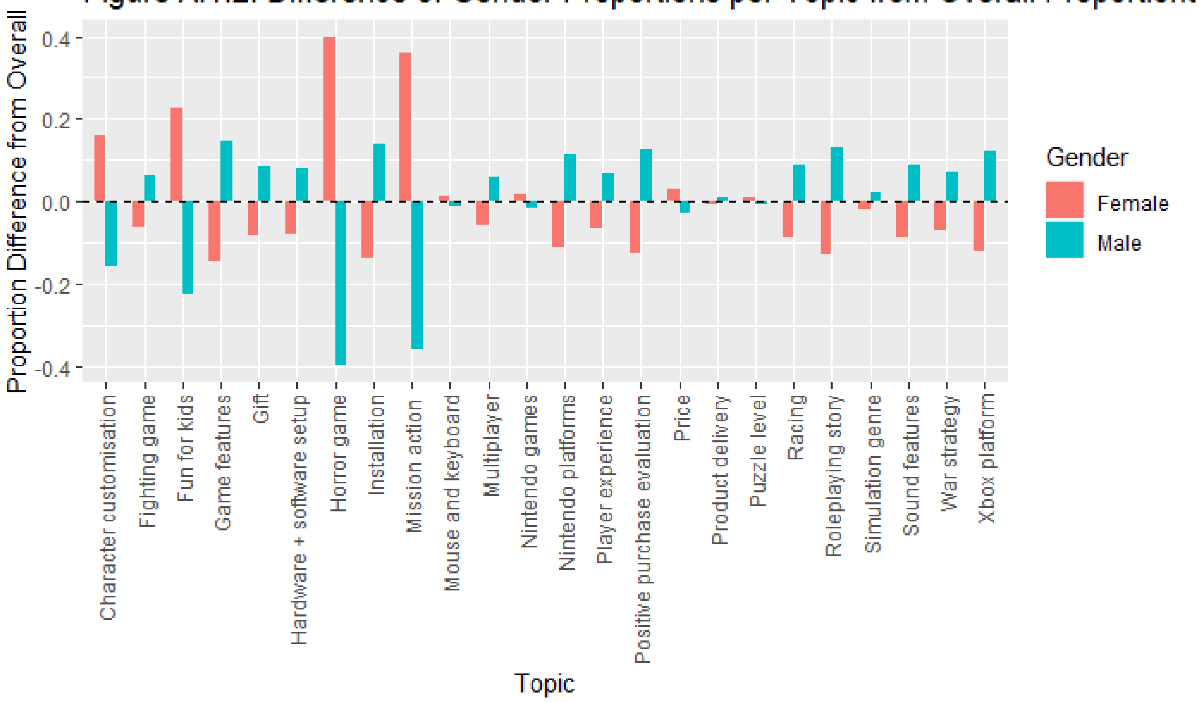


Figure A.2. Top 20 Important Features, model 2

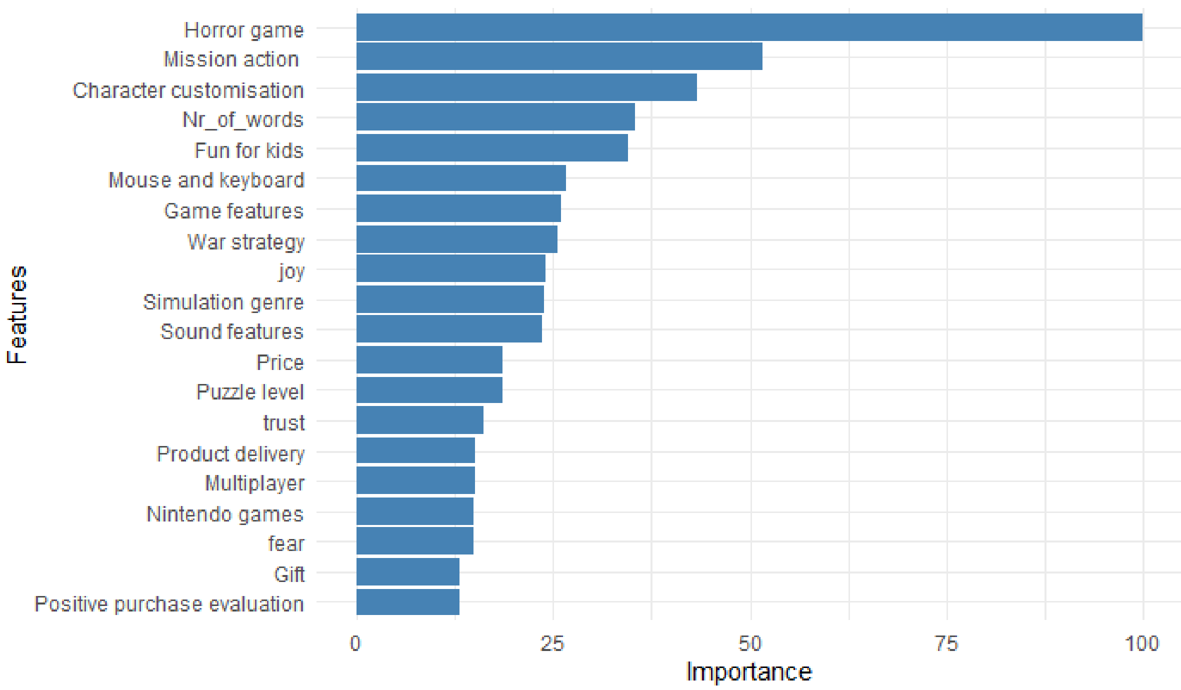


Figure A.3. Top 20 Important Features, model 3

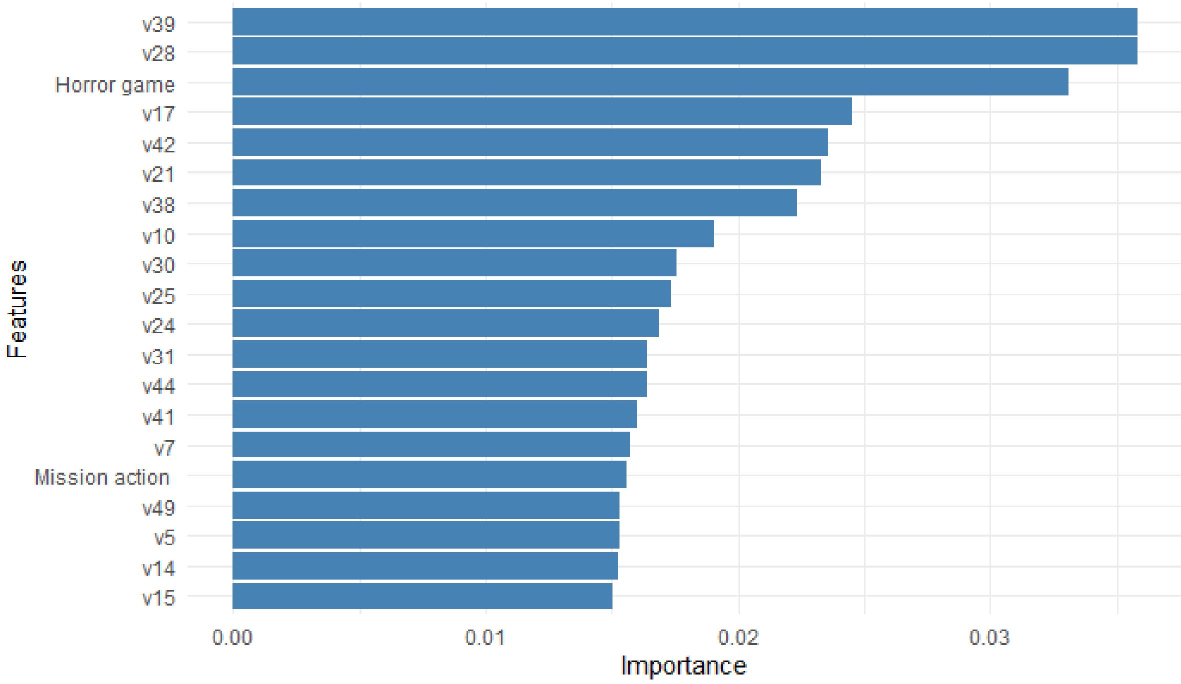


Figure A.4. Top 20 Important Features, model 4

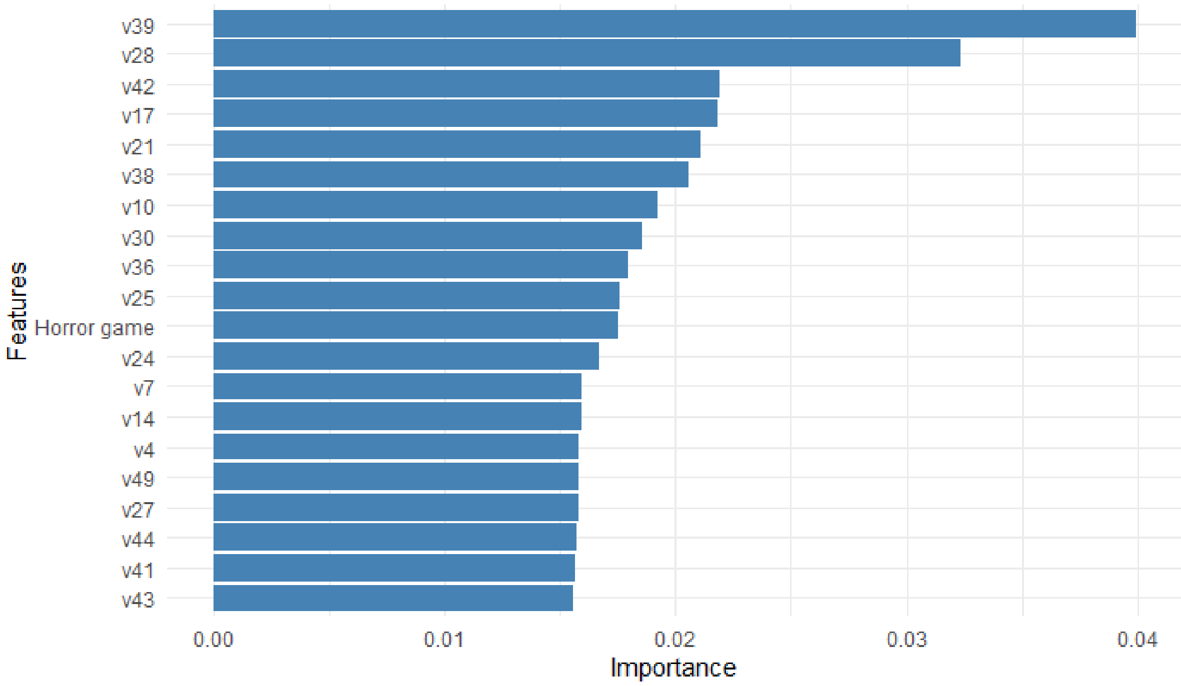


Figure A.5. Top 20 Important Features, model 5

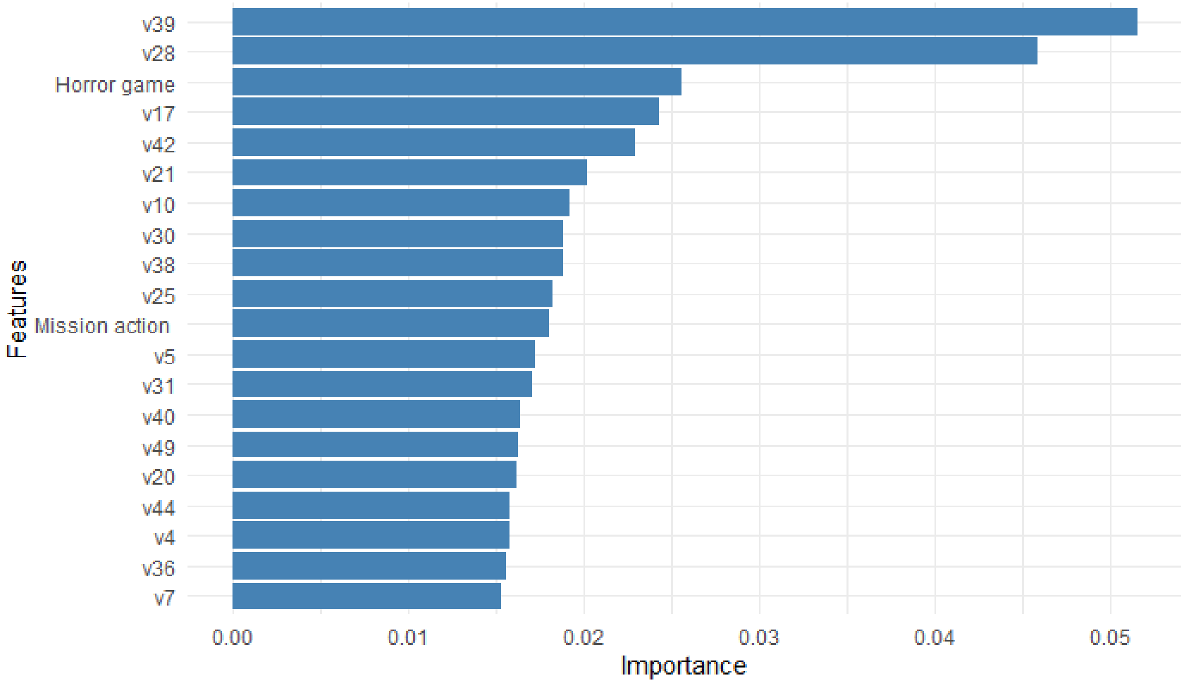


Figure A.6. Partial Dependency for Mission action topic, model 2

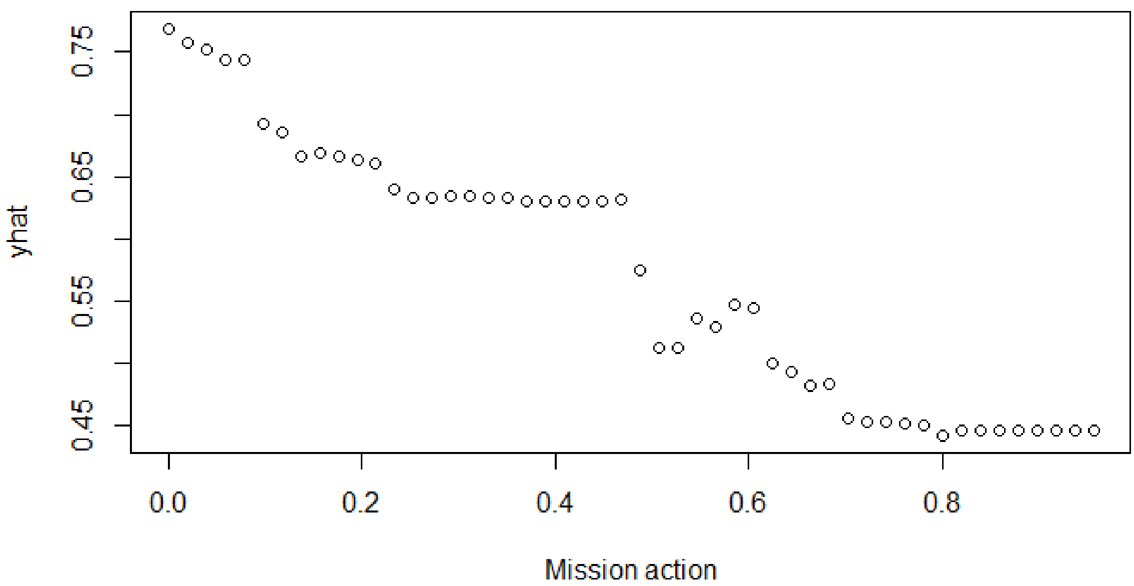


Figure A.7. Partial Dependency for Mission action topic, model 5

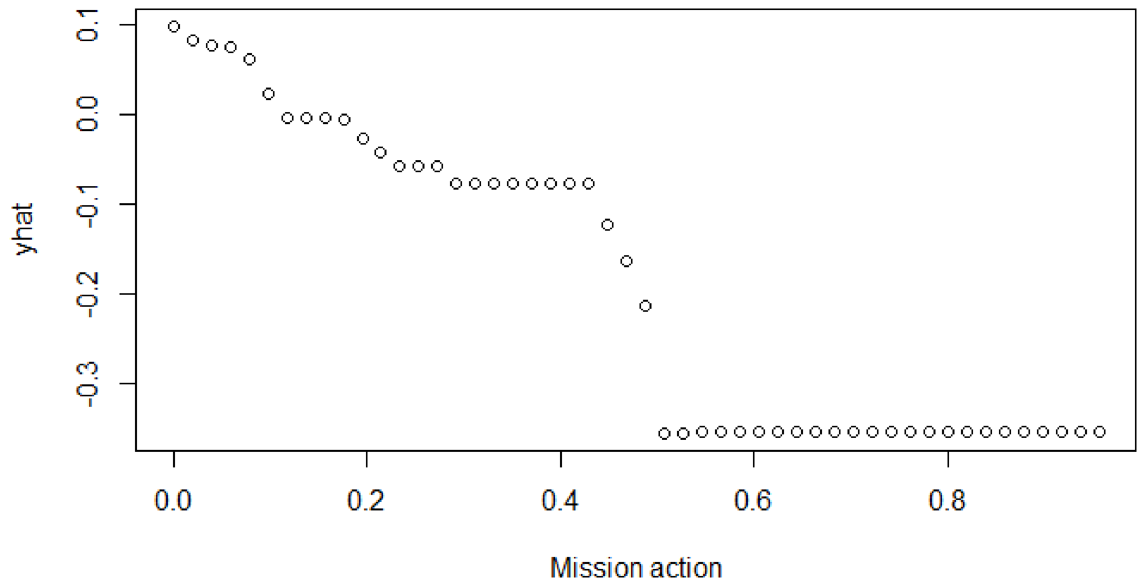


Figure A.8. Partial Dependency for horror genre topic, model 2

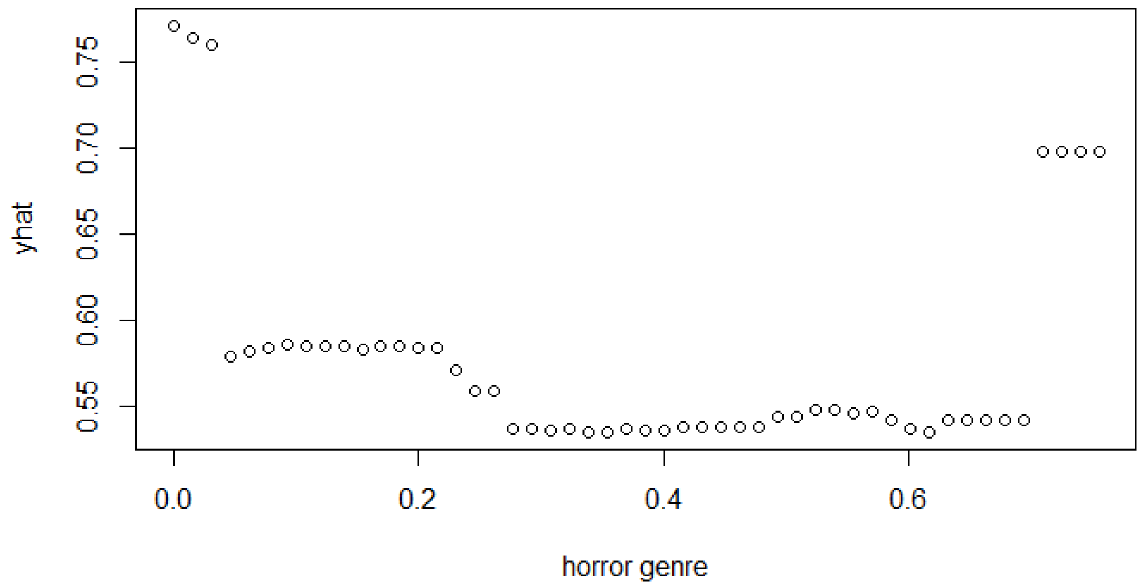


Figure A.9. Partial Dependency for horror genre topic, model 5

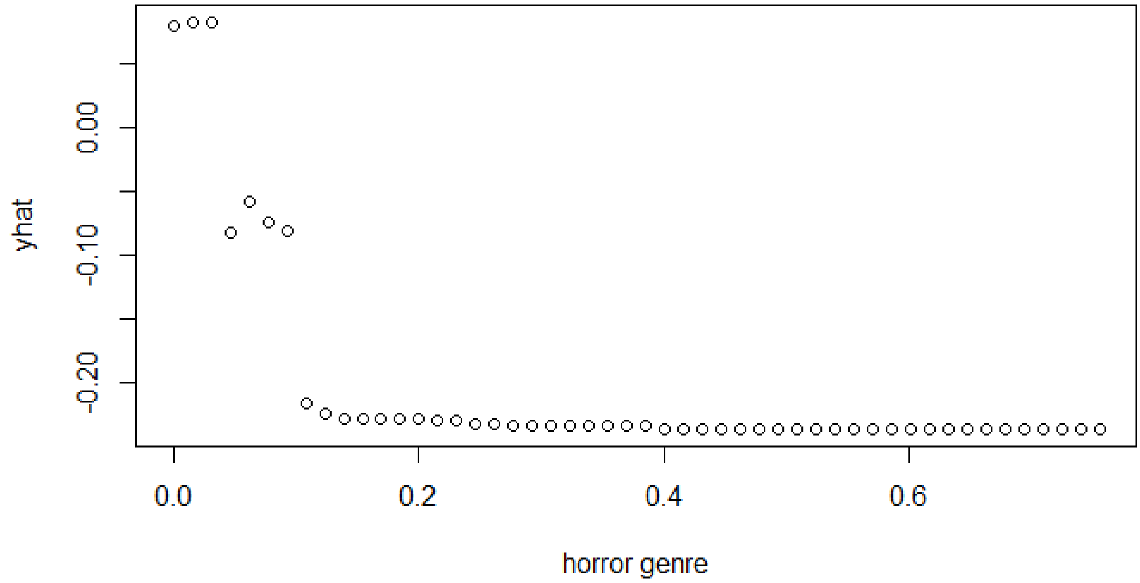


Figure A.10. Partial Dependency for Multiplayer topic, model 2

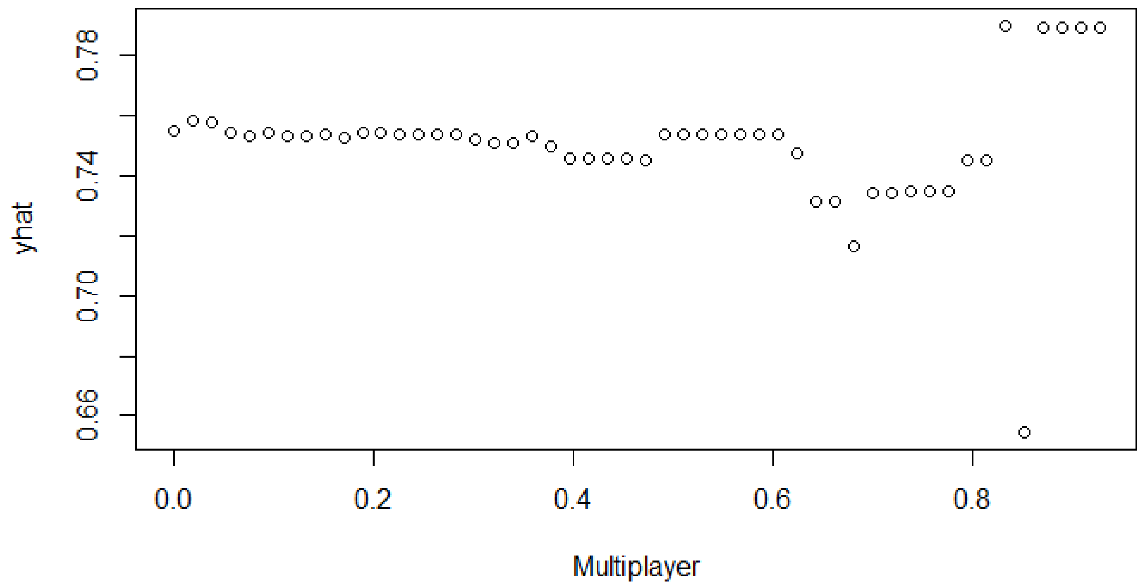


Figure A.11. Partial Dependency for Multiplayer topic, model 5

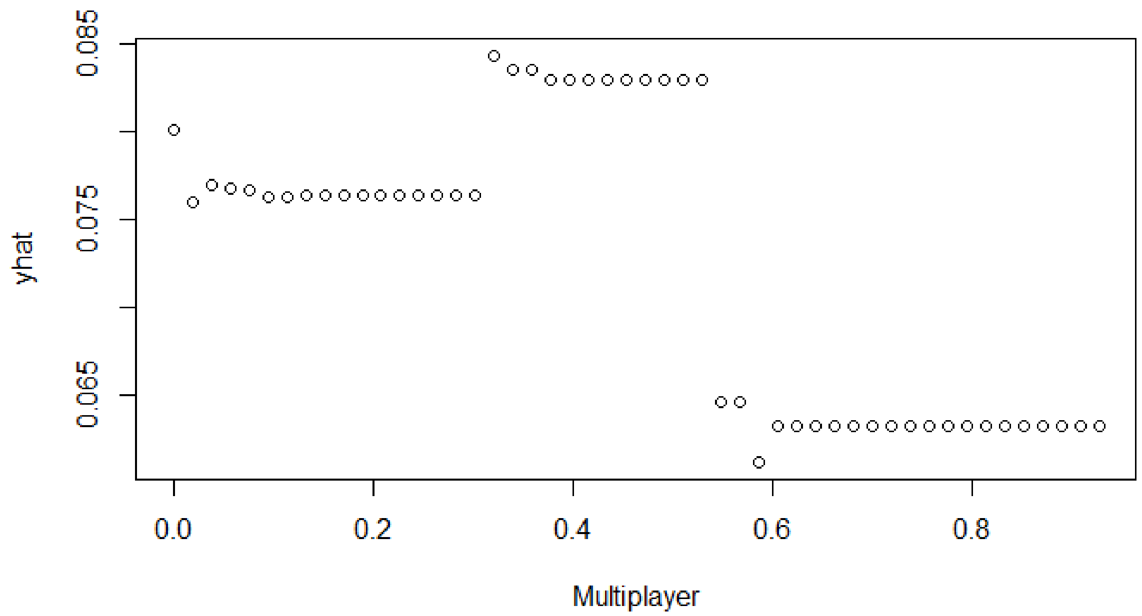


Figure A.12. Partial Dependency for Graphics topic, model 2

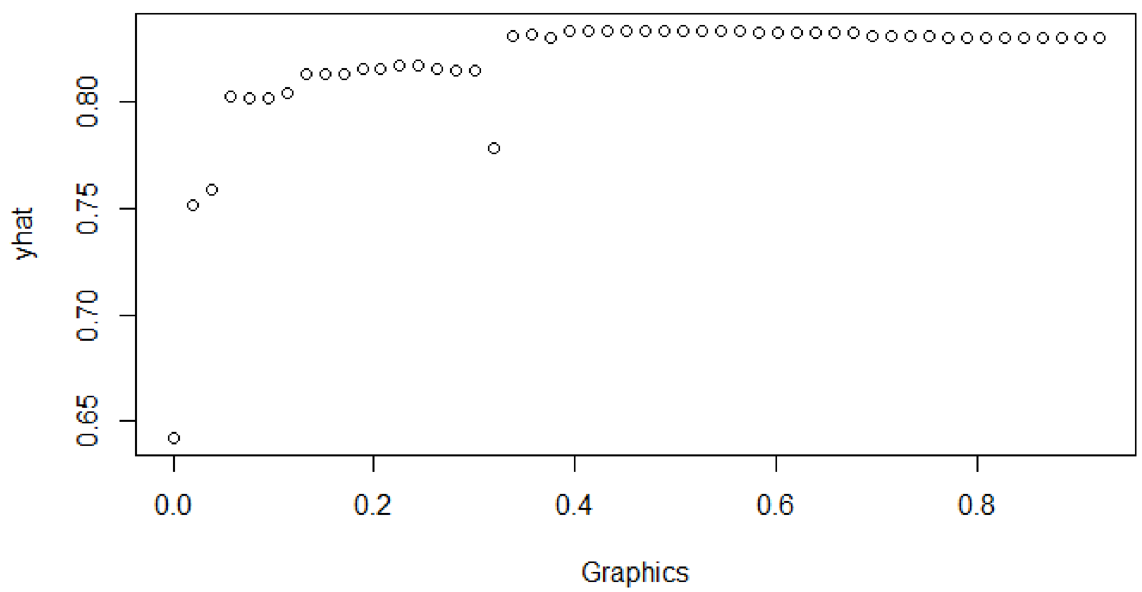


Figure A.13. Partial Dependency for Graphics topic, model 5

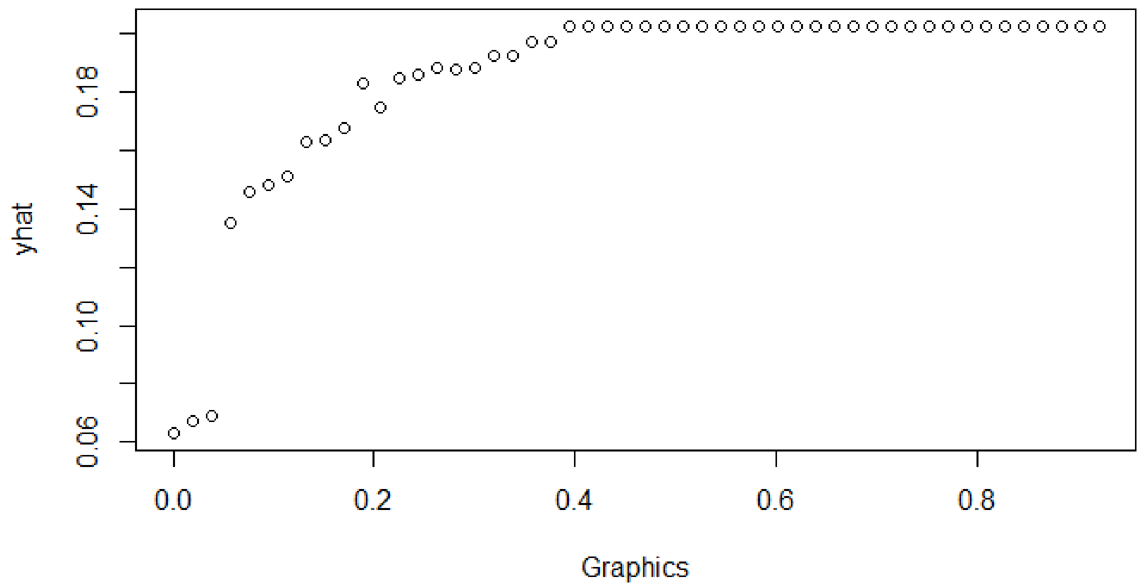


Figure A.14. Partial Dependency for Sound topic, model 2

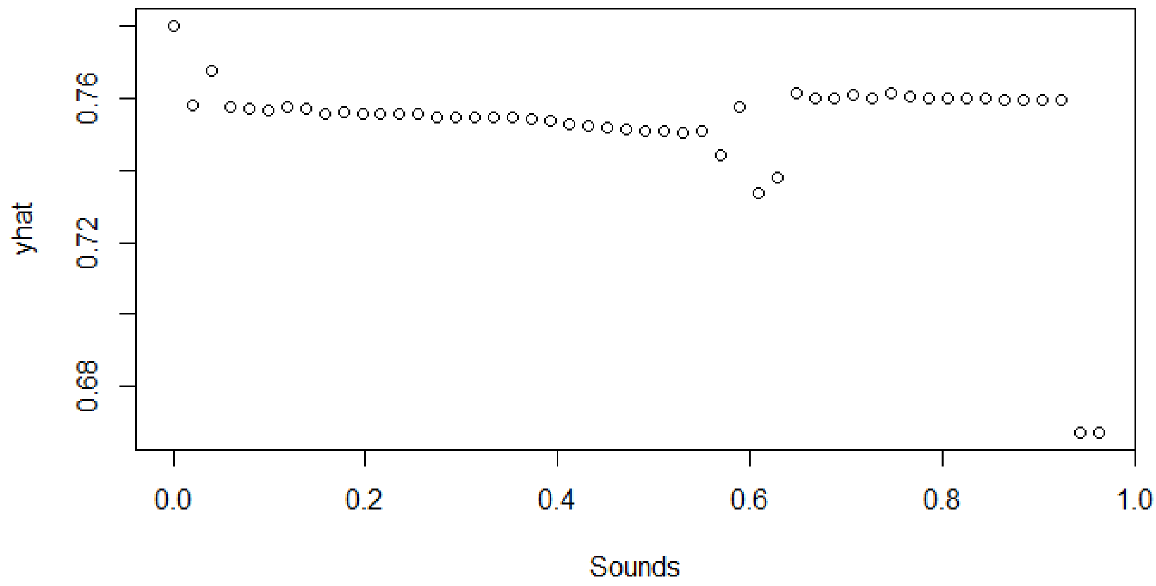


Figure A.15. Partial Dependency for Sound topic, model 5

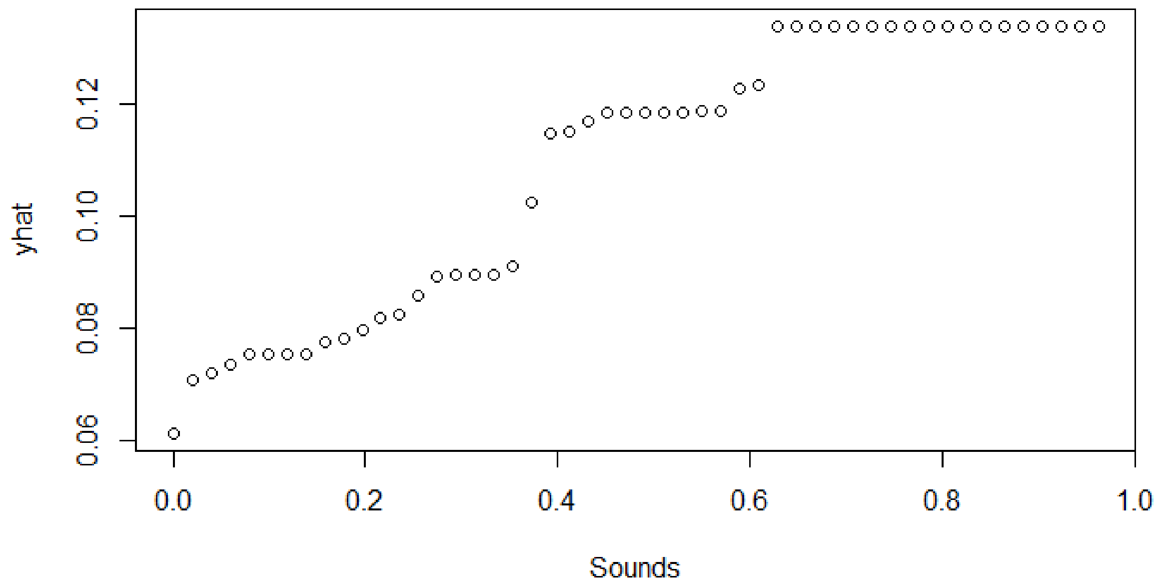


Figure A.16. Partial Dependency for Racing topic, model 2

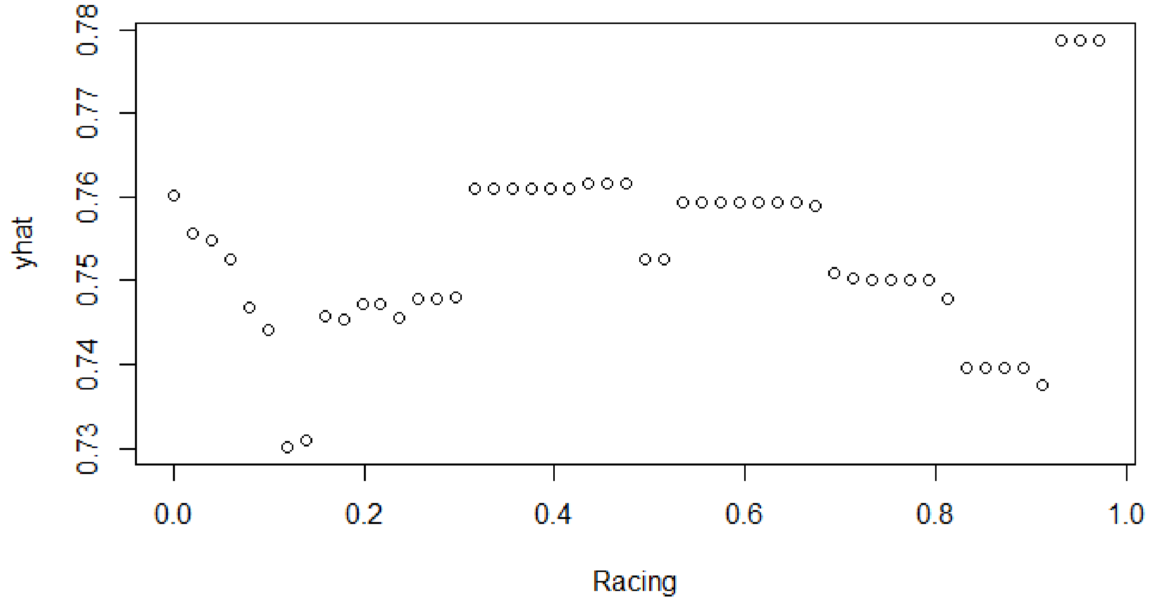


Figure A.17. Partial Dependency for Racing topic, model 5

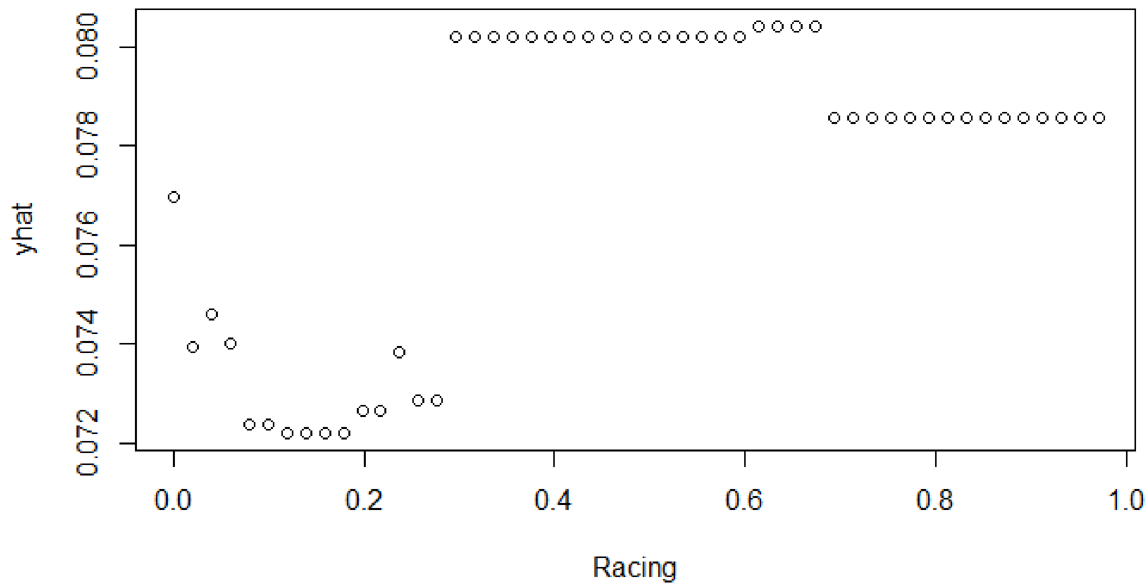


Figure A.18. Partial Dependency for War strategy topic, model 2

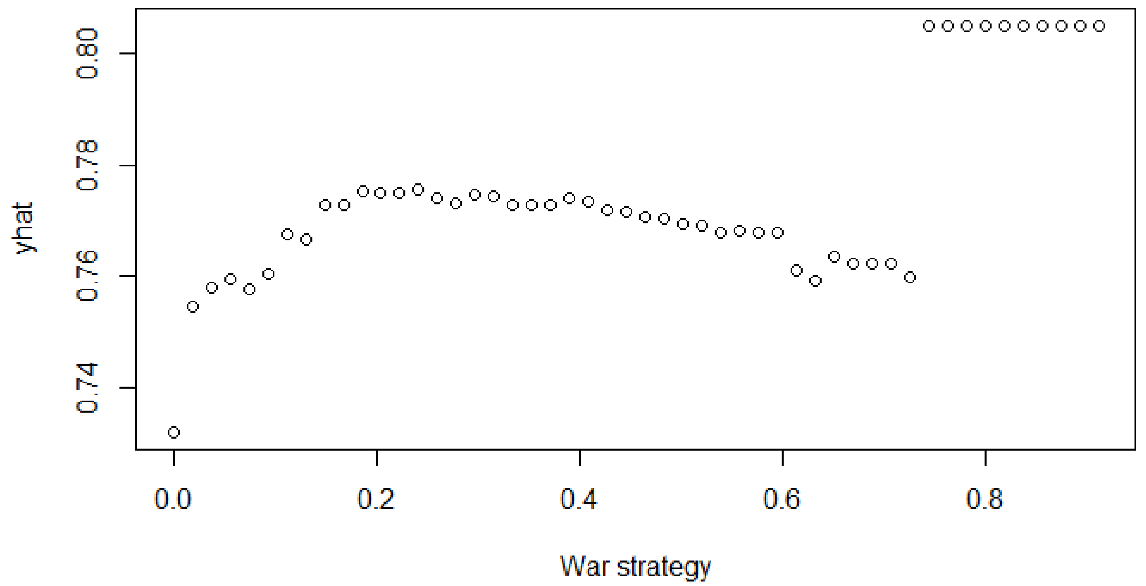


Figure A.19. Partial Dependency for War strategy topic, model 5

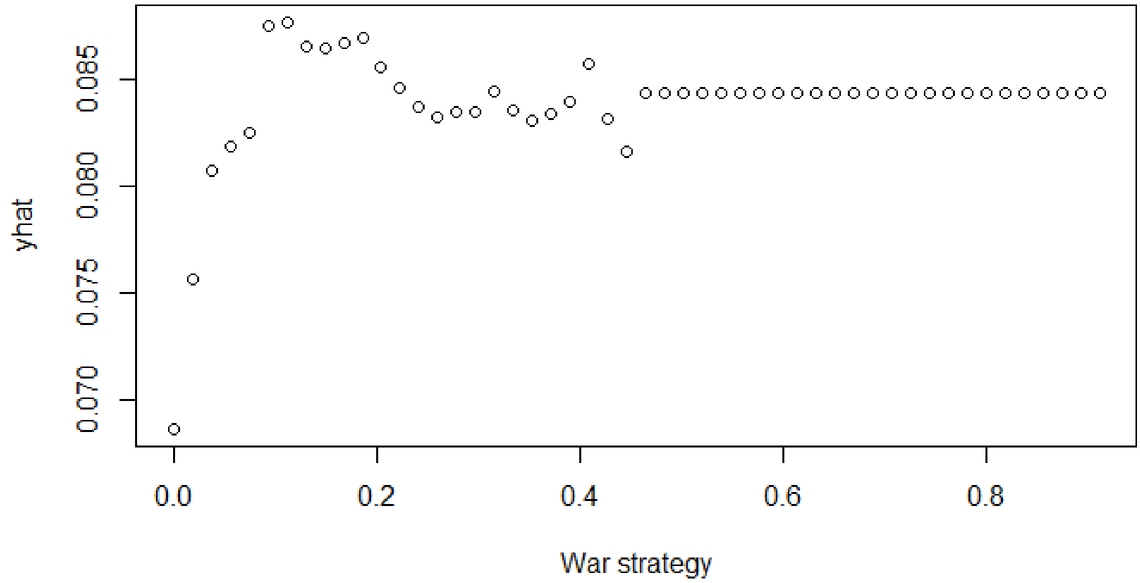


Figure A.20. Partial Dependency for Simulation strategy topic, model 2

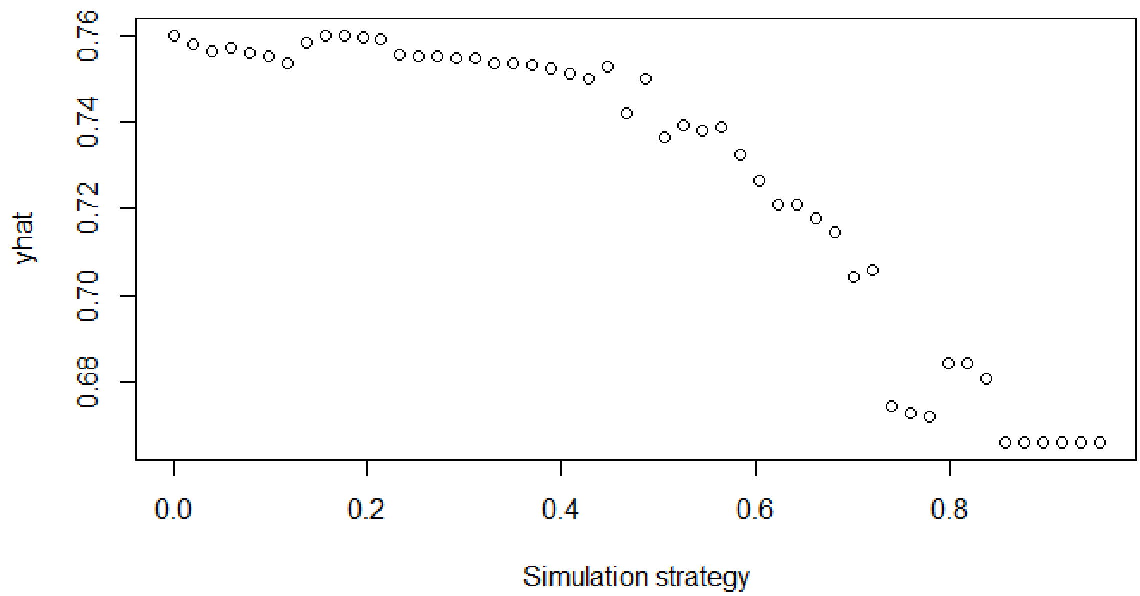


Figure A.21. Partial Dependency for Simulation strategy topic, model 5

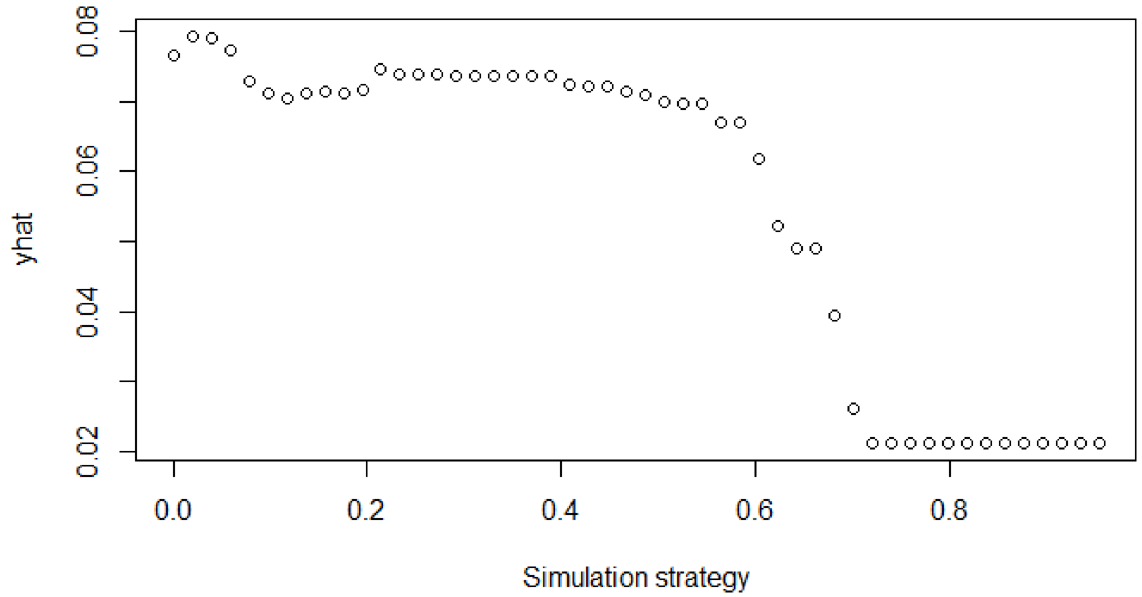


Figure A.22. Distribution of Votes

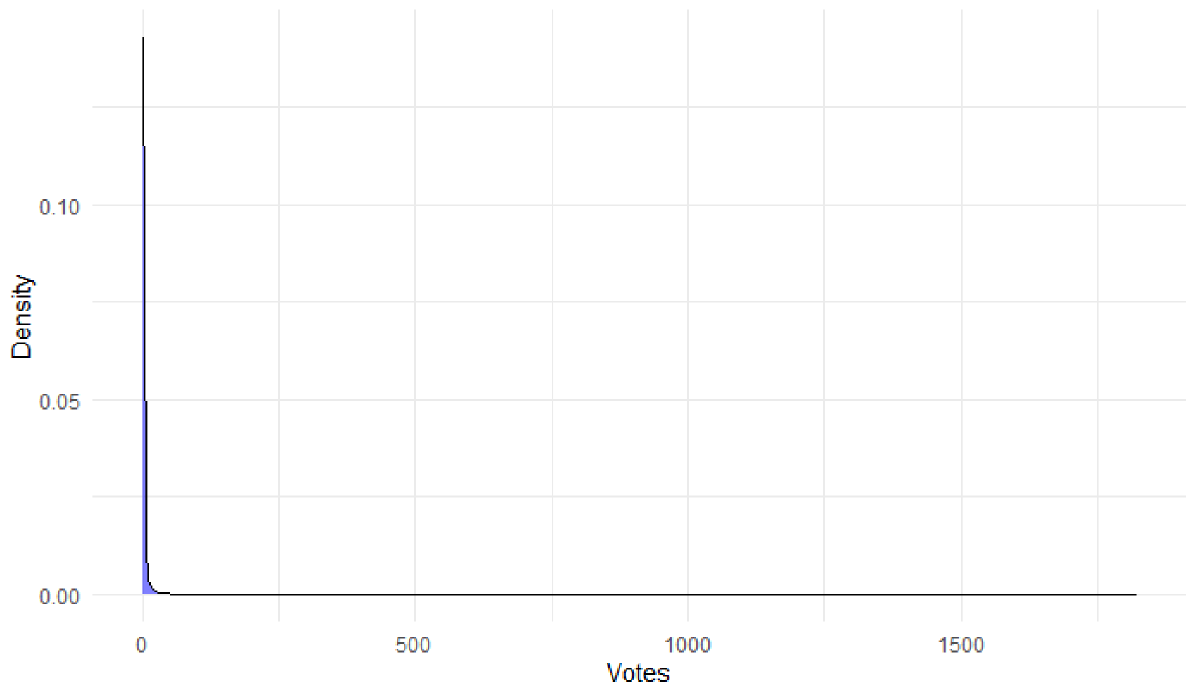


Figure A.23. Distribution of Votes after trimming

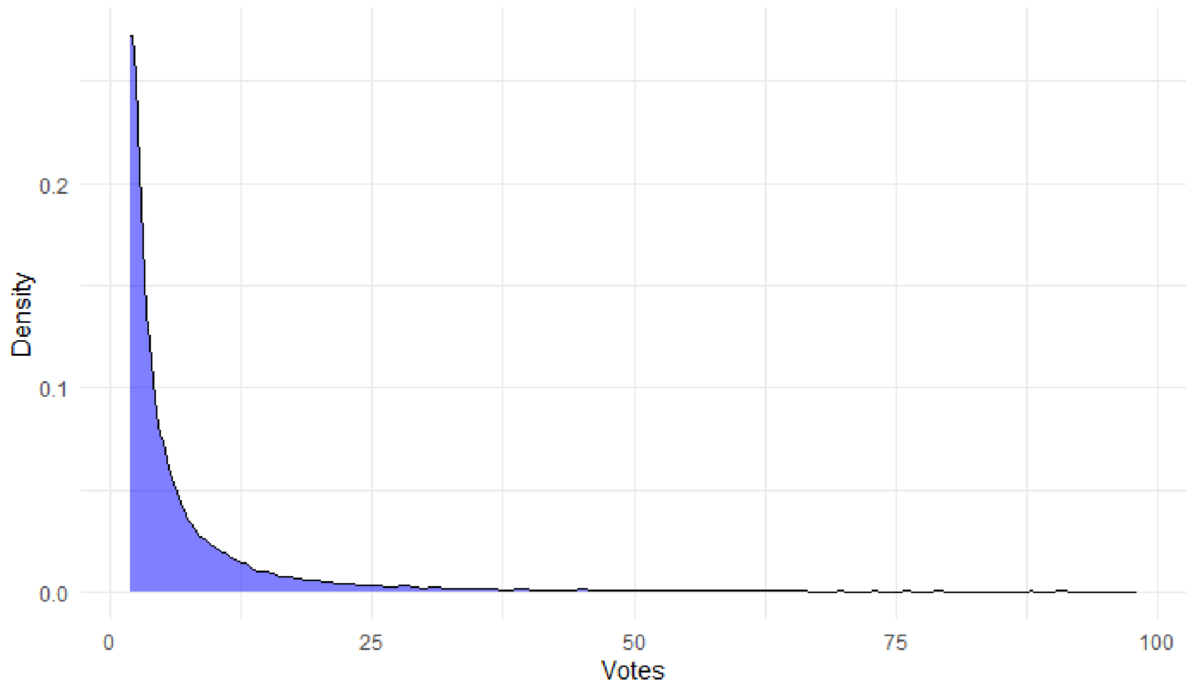


Figure A.24 Observed vs Predicted Values for Helpfulness model 1

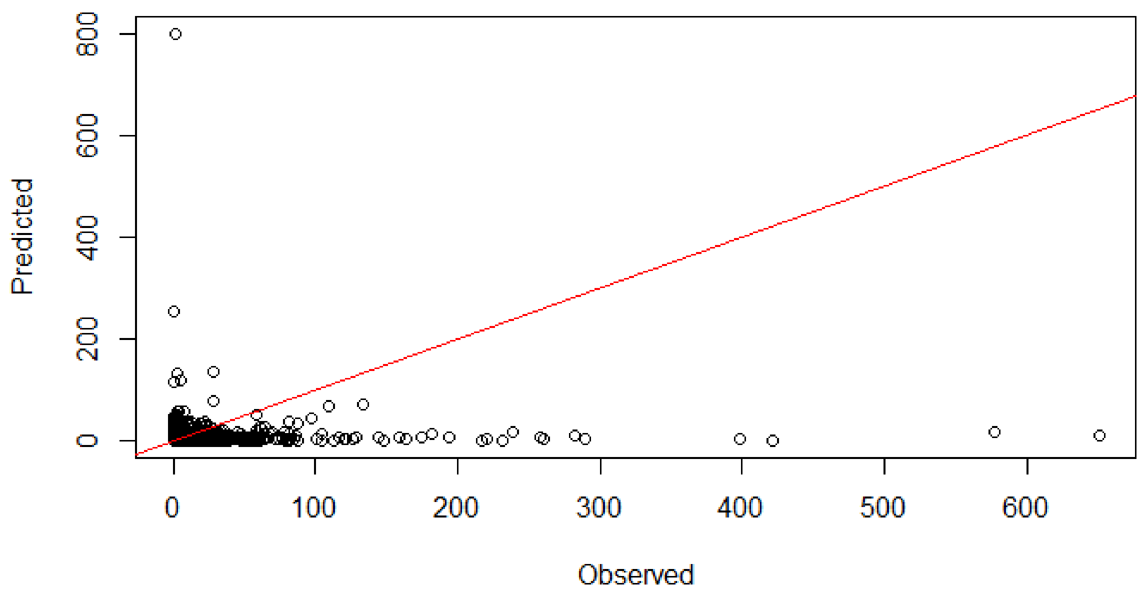


Figure A.25 Observed vs Predicted Values for Helpfulness model 2

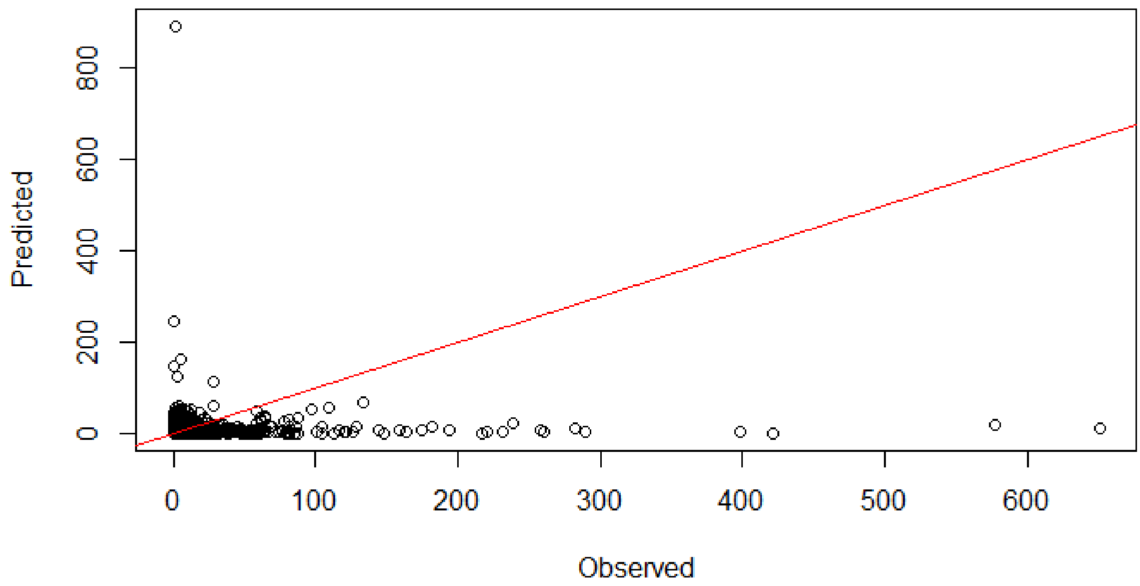


Figure A.26 Observed vs Predicted Values for Helpfulness model 3

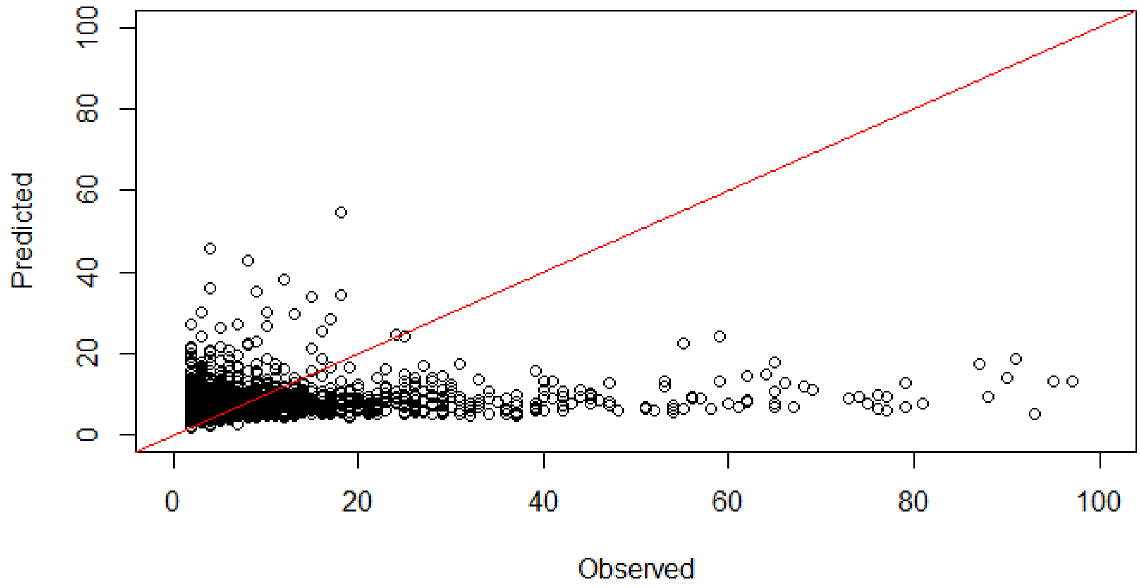


Table A.1. Regression results for helpfulness model 1

	Dependent variable: Votes
gendermale	-0.05 (-0.10, 0.01)
Nr_of_words	-0.0002*** (-0.0003, -0.0001)
Rating	-0.18*** (-0.19, -0.17)
anger	-0.02*** (-0.02, -0.02)
anticipation	0.05*** (0.05, 0.06)
disgust	0.01*** (0.01, 0.02)
fear	-0.05*** (-0.05, -0.04)
joy	-0.05*** (-0.05, -0.05)
sadness	0.04*** (0.04, 0.05)
surprise	-0.02*** (-0.02, -0.01)
trust	0.002 (-0.001, 0.005)
negative	-0.01*** (-0.01, -0.005)
positive	0.07*** (0.07, 0.07)
Horror game	-31.53*** (-32.64, -30.42)
Mission action	0.83*** (0.75, 0.91)
Multiplayer	0.83*** (0.72, 0.94)
Simulation	-0.28*** (-0.36, -0.20)
War strategy	0.66*** (0.55, 0.76)
Racing	-0.49*** (-0.65, -0.34)
Graphics	-0.42*** (-0.59, -0.24)
Sound	0.72*** (0.64, 0.79)
gendermale:Rating	-0.01** (-0.02, -0.003)
gendermale:negative	-0.01*** (-0.01, -0.01)
gendermale:Horror game	-19.75*** (-21.14, -18.35)
gendermale:Mission action	0.55*** (0.45, 0.66)
gendermale:Multiplayer	0.47*** (0.35, 0.59)
gendermale:Simulation	0.45*** (0.36, 0.54)
gendermale:War strategy	1.16*** (1.05, 1.27)
gendermale:Racing	0.31*** (0.14, 0.47)
gendermale:Graphics	0.38*** (0.19, 0.57)
gendermale:Sound	-0.49*** (-0.58, -0.41)
Constant	1.22*** (1.17, 1.27)
Observations	68,103
Log Likelihood	-242,979.70
Akaike Inf. Crit.	486,023.50
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.2 Helpfulness votes model 2 regression results

	Dependent variable: Votes

Rating	-0.17*** (-0.17, -0.17)
Nr_of_words	-0.001*** (-0.001, -0.001)
anger	-0.02*** (-0.03, -0.02)
anticipation	0.05*** (0.05, 0.06)
disgust	0.02*** (0.02, 0.02)
fear	-0.04*** (-0.04, -0.03)
joy	-0.05*** (-0.05, -0.04)
sadness	0.04*** (0.04, 0.04)
surprise	-0.02*** (-0.02, -0.02)
trust	0.004*** (0.001, 0.01)
negative	-0.01*** (-0.01, -0.01)
positive	0.07*** (0.06, 0.07)
Fighting game	-1.86*** (-1.99, -1.73)
Product delivery	-0.14*** (-0.22, -0.06)
Horror game	-32.20*** (-33.09, -31.32)
Puzzle level	-1.44*** (-1.54, -1.34)
Fun for kids	0.30*** (0.22, 0.39)
Game features	-0.75*** (-0.84, -0.66)
Nintendo platforms	-0.59*** (-0.68, -0.49)
Mission action	0.51*** (0.43, 0.59)
Roleplaying story	-0.92*** (-1.02, -0.83)
Sound features	-0.21*** (-0.28, -0.14)
Xbox platform	-0.72*** (-0.80, -0.63)
Positive purchase evaluation	-1.83*** (-1.95, -1.70)
Hardware + software setup	-0.24*** (-0.33, -0.15)
Mouse and keyboard	-1.10*** (-1.21, -1.00)
Player experience	-0.03 (-0.11, 0.06)
Gift	0.41*** (0.32, 0.50)
Simulation genre	-0.55*** (-0.62, -0.49)
Multiplayer	0.62*** (0.54, 0.69)
War strategy	1.04*** (0.97, 1.11)
Installation	-0.12 (-0.29, 0.05)
Racing	-0.88*** (-0.96, -0.80)
Price	-12.53*** (-13.16, -11.91)
Character customisation	-1.39*** (-1.48, -1.29)
Nintendo games	
male_dominance_score	
female_dominance_score	
gender_recoded	-0.06*** (-0.09, -0.04)
female_content_congruence	0.03 (-0.03, 0.08)
male_content_congruence	-0.07*** (-0.12, -0.02)
Constant	1.79*** (1.73, 1.85)
Observations	68,103
Log Likelihood	-238,635.10
Akaike Inf. Crit.	477,348.30
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.3 Helpfulness votes model 3 regression results

	Dependent variable: Votes
Rating	-0.06*** (-0.06, -0.05)
Nr_of_words	-0.0000 (-0.0002, 0.0001)
anger	-0.01*** (-0.01, -0.003)
anticipation	0.02*** (0.02, 0.03)
disgust	-0.001 (-0.01, 0.004)
fear	0.02*** (0.01, 0.02)
joy	-0.01*** (-0.01, -0.01)
sadness	-0.01*** (-0.02, -0.01)
surprise	-0.01*** (-0.01, -0.005)
trust	-0.01*** (-0.01, -0.003)
negative	-0.01*** (-0.01, -0.01)
positive	0.03*** (0.03, 0.03)
Fighting game	-0.70*** (-0.83, -0.57)
Product delivery	0.30*** (0.20, 0.40)
Horror game	-0.87*** (-1.12, -0.63)
Puzzle level	-0.65*** (-0.76, -0.54)
Fun for kids	0.06 (-0.05, 0.18)
Game features	-0.29*** (-0.39, -0.19)
Nintendo platforms	-0.30*** (-0.42, -0.19)
Mission action	-0.08 (-0.18, 0.02)
Roleplaying story	-0.17*** (-0.28, -0.06)
Sound features	-0.38*** (-0.47, -0.29)
Xbox platform	-0.38*** (-0.48, -0.28)
Positive purchase evaluation	-0.07 (-0.20, 0.06)
Hardware + software setup	-0.14** (-0.24, -0.03)
Mouse and keyboard	-0.51*** (-0.62, -0.39)
Player experience	-0.17*** (-0.28, -0.07)
Gift	0.38*** (0.28, 0.49)
Simulation genre	-0.21*** (-0.30, -0.13)
Multiplayer	0.21*** (0.12, 0.30)
War strategy	0.35*** (0.27, 0.44)
Installation	-0.84*** (-1.04, -0.64)
Racing	-0.27*** (-0.38, -0.16)
Price	-1.80*** (-2.05, -1.54)
Character customisation	-0.30*** (-0.40, -0.20)
Nintendo games	
male_dominance_score	
female_dominance_score	
gender_recoded	-0.09*** (-0.12, -0.07)
female_content_congruence	0.14*** (0.07, 0.21)
male_content_congruence	0.10*** (0.04, 0.15)
Constant	2.14*** (2.06, 2.22)
Observations	12,214
Log Likelihood	-66,351.76
Akaike Inf. Crit.	132,781.50

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.4 Helpfulness votes model 4 regression results

	Dependent variable: Helpful
gendermale	0.70*** (0.48, 0.92)
Nr_of_words	0.01*** (0.005, 0.01)
Rating	-0.25*** (-0.29, -0.21)
anger	-0.003 (-0.03, 0.02)
anticipation	0.08*** (0.06, 0.09)
disgust	0.01 (-0.02, 0.03)
fear	-0.06*** (-0.08, -0.04)
joy	0.01 (-0.01, 0.03)
sadness	0.01 (-0.01, 0.03)
surprise	-0.03*** (-0.06, -0.01)
trust	-0.01 (-0.03, 0.01)
negative	0.08*** (0.06, 0.11)
positive	0.03*** (0.02, 0.04)
Horror game	-4.12*** (-5.20, -3.04)
Mission action	0.99*** (0.60, 1.38)
Multiplayer	0.85*** (0.36, 1.33)
Simulation	0.61*** (0.26, 0.95)
War strategy	2.36*** (1.91, 2.80)
Racing	-0.95*** (-1.57, -0.33)
Graphics	-0.06 (-0.79, 0.67)
Sound	0.74*** (0.41, 1.08)
gendermale:Rating	-0.04** (-0.08, -0.001)
gendermale:negative	-0.10*** (-0.11, -0.08)
gendermale:Horror game	-21.90*** (-24.47, -19.33)
gendermale:Mission action	0.17 (-0.37, 0.71)
gendermale:Multiplayer	0.38 (-0.15, 0.92)
gendermale:Simulation	0.15 (-0.24, 0.53)
gendermale:War strategy	-0.66*** (-1.16, -0.17)
gendermale:Racing	-0.09 (-0.77, 0.58)
gendermale:Graphics	0.80** (0.03, 1.57)
gendermale:Sound	0.03 (-0.33, 0.40)
Constant	-1.42*** (-1.61, -1.23)
Observations	68,103
Log Likelihood	-25,662.76
Akaike Inf. Crit.	51,389.53
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.5 Helpfulness votes model 5 regression results

	Dependent variable:

	Helpful
Rating	-0.27*** (-0.29, -0.26)
Nr_of_words	0.01*** (0.01, 0.01)
anger	-0.003 (-0.03, 0.02)
anticipation	0.08*** (0.06, 0.10)
disgust	-0.001 (-0.03, 0.03)
fear	-0.06*** (-0.08, -0.04)
joy	0.003 (-0.02, 0.02)
sadness	0.01 (-0.01, 0.03)
surprise	-0.03*** (-0.06, -0.01)
trust	-0.0004 (-0.02, 0.02)
negative	0.003 (-0.01, 0.02)
positive	0.03*** (0.01, 0.04)
Fighting game	-0.78*** (-1.23, -0.34)
Product delivery	-0.50*** (-0.84, -0.16)
Horror game	-8.94*** (-10.21, -7.68)
Puzzle level	-0.04 (-0.40, 0.33)
Fun for kids	-0.91*** (-1.35, -0.46)
Game features	0.19 (-0.15, 0.53)
Nintendo platforms	-0.41** (-0.80, -0.02)
Mission action	0.56*** (0.20, 0.92)
Roleplaying story	-0.50*** (-0.85, -0.15)
Sound features	0.44*** (0.16, 0.72)
Xbox platform	-0.06 (-0.41, 0.29)
Positive purchase evaluation	-1.50*** (-1.95, -1.04)
Hardware + software setup	0.76*** (0.41, 1.12)
Mouse and keyboard	0.42** (0.08, 0.77)
Player experience	-0.99*** (-1.36, -0.62)
Gift	0.42** (0.07, 0.78)
Simulation genre	0.40*** (0.12, 0.68)
Multiplayer	0.82*** (0.51, 1.14)
War strategy	1.48*** (1.18, 1.79)
Installation	-2.34*** (-3.11, -1.57)
Racing	-1.39*** (-1.74, -1.05)
Price	-6.51*** (-7.41, -5.61)
Character customisation	0.18 (-0.15, 0.50)
Nintendo games	
male_dominance_score	
female_dominance_score	
gender_recoded	-0.10** (-0.20, -0.001)
female_content_congruence	-0.38*** (-0.63, -0.14)
male_content_congruence	0.21** (0.01, 0.41)
Constant	-0.63*** (-0.88, -0.37)
Observations	68,103
Log Likelihood	-25,448.69
Akaike Inf. Crit.	50,975.38
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix B - Other figures and tables

Table B.1. Sanity Check, t-test on sample means

Variable	p-value
Rating	0.51
Nr of words	0.01
Anger	0.00
Anticipation	0.01
Disgust	0.02
Fear	0.00
Joy	0.01
Sadness	0.00
Surprise	0.02
Trust	0.00
Negative	0.00
Positive	0.00
Fighting game mode topic	0.12
Product delivery topic	0.77
Horror game topic	0.02
Puzzle level topic	0.50
Fun for kids topic	0.04
Game Features (graphics, story, gameplay) topic	0.98
Nintendo platforms topic	0.04
Mission-based action topic	0.76
Roleplaying story game topic	0.69
Sound features topic	0.43
Xbox platform topic	0.20

Positive purchase evaluation topic	0.01
Setup, hardware (screen, buttons) + software (system) topic	0.75
Mouse and keyboard topic	0.66
Player experience topic	0.99
Gift for son topic	0.50
Simulation genre (e.g. city expansion) topic	0.31
Multiplayer topic	0.49
War strategy genre topic	0.84
Installation (platform & version) topic	0.96
Racing genre topic	0.01
Price topic	0.77
Character design/customisation topic	0.25
Nintendo games (e.g. Mario) topic	0.05

These p-values are obtained by performing a t-test on the difference in means between a random sample of 24000 observations from the overall original dataset (before discarding samples) and a random sample of 24000 observations from the discarded dataset.

Table B.2. LDA topics and their names based on term association

Topic 1	Fighting game mode
Topic 2	Product delivery
Topic 3	Horror game
Topic 4	Puzzle level
Topic 5	Fun for kids
Topic 6	Game Features (graphics, story, gameplay)
Topic 7	Nintendo platforms
Topic 8	Mission-based action
Topic 9	Roleplaying story game
Topic 10	Sound features

Topic 11	Xbox platform
Topic 12	Positive purchase evaluation
Topic 13	Setup, hardware (screen, buttons) + software (system)
Topic 14	Mouse and keyboard
Topic 15	Player experience
Topic 16	Gift for son
Topic 17	Simulation genre (e.g. city expansion)
Topic 18	Multiplayer
Topic 19	War strategy genre
Topic 20	Installation (platform & version)
Topic 21	Racing genre
Topic 22	Price
Topic 23	Character design/customisation
Topic 24	Nintendo games (e.g. Mario)

These topic names are based on personal interpretations of the associated terms and are not necessarily absolute.

Table B.3. Hyperparameters used in model training

Learning rate	0.01, 0.1, 0.3
Max depth	4, 6, 8, 10

These hyperparameter values and their possible combinations were tested when training the XGBoost models to find the best performing model. Parameters not included in this table were set to their default values as per the xgboost package.

Appendix C - Files

All files used can be found in this google drive:
https://drive.google.com/drive/folders/1bqNdeykPPonYi-dQrP7KizVzonHLlJdt?usp=sharing