Ca' Foscari
University
of Venice

Master's Degree programme

in Language Sciences

Final Thesis

Youtube emotions and Ratings on Amazon

An in-depth analysis

Supervisor
Ch. Prof. Gianluca Lebani

Assistant supervisor
Ch. Prof. Francesca Santulli

Graduand
Fatemeh Ardestani
Matriculation Number 893295

Academic Year
2023 / 2024

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Gianluca Lebani, for his guidance and support throughout this journey. His patience have been instrumental in shaping this thesis.

I would also like to extend my gratitude to Professor Francesca Santulli for her supervision and the effort she has put in to bring this work to fruition.

Last but not the least, I would like to thank my family for their support, encouragement, and love throughout this journey. This achievement would not have been possible without their understanding, patience, and sacrifices. Thank you for always standing by my side and inspiring me to pursue my dreams.

Abstract

This thesis examines the relationship between product ratings on Amazon and the emotions expressed in transcribed YouTube product reviews. The primary objectives were to determine whether the sentiments in YouTube reviews align with the product ratings on Amazon, and to assess the connection between viewer comments on those reviews and Amazon ratings. Quantitative methods, including linear models, were employed to explore the associations between emotional content and ratings. A notable correlation was discovered in our research between Amazon product ratings and the sentiments expressed in YouTube reviews. Additionally, a similar correlation was identified between the emotions conveyed in the comments section of YouTube reviews and Amazon ratings. However, the study found that Amazon ratings tend to skew toward higher values, typically above 4, which may influence the strength of this correlation. Furthermore, a noticeable divergence in the emotional characteristics of the review transcriptions and the comments suggests that different emotional dynamics are present in the videos and the corresponding viewer responses. The significance of this study lies in its potential applications. The findings could be used in the development of automated systems that evaluate YouTube reviews, providing consumers with insights into a product's quality without the need to watch the entire video. Additionally, by identifying discrepancies between user-generated content and Amazon ratings, this approach could help detect false or deceptive ratings. While the study found a correlation between ratings and emo-

tions, ratings should not be predicted solely based on emotions. Since this research focused exclusively on emotions derived from transcriptions, future studies could perform a multimodal analysis that incorporates visual and prosodic data alongside textual information from reviews. These findings pave the way for more advanced multimodal analysis methods and highlight the importance of incorporating emotional analysis into product rating evaluations.

Keywords Emotion Analysis, Review Rating, Online Review

Contents

Chapter 1

Introduction

The rise of online shopping has fundamentally transformed consumer behavior and the way purchasing decisions are made. Consumers now have unprecedented access to a vast range of products, offering them choices that go beyond the constraints of local availability. While this access has changed the shopping experiences for the better in many ways, it has also presented challenges in evaluating product quality. The sheer volume of new products entering the market, often sold by unfamiliar vendors, has created a strong reliance on online reviews. These reviews, available in various formats such as text, video, and aggregated ratings, have become an essential tool for consumers seeking to verify the legitimacy of marketing claims and to gain insights into the actual performance of products.

Two prominent platforms, Amazon and YouTube, have emerged as key players in the realm of online reviews. Amazon, the global e-commerce giant, offers user-generated star ratings and written reviews, which provide quantitative and qualitative measures of product satisfaction. On the other hand, YouTube serves as a popular platform for video-based product reviews, where influencers and everyday users share detailed experiences and technical insights about a product's features, functionality, and performance. Both platforms, while valuable, have distinct modes of influencing consumer behavior and are susceptible to manipulation. On

YouTube, paid promotions can skew the authenticity of reviews, while on Amazon, the problem of fake ratings has been well documented.

Furthermore, both platforms face limitations in review availability. For newly launched or lesser-known products, the absence of sufficient reviews or ratings can leave consumers uncertain about their choices. This research aims to explore the relationship between the emotions expressed in YouTube reviews and the star ratings on Amazon, with the goal of uncovering how emotions might correlate with product evaluations across these platforms. BERT models (Bidirectional Encoder Representations from Transformers), based on the transformers[1] architecture, are pretrained models that learn contextual relationships between words in a sentence by training on large amounts of text data. RoBERTa[2] (Robustly Optimized BERT Approach) as the name suggests, is an improved BERT with optimized training process which performs better than BERT[3]. By employing a BERT-based model[4], specifically Facebook's RoBERTa[5] fine-tuned on Google's GoEmotions[6] dataset[7], we perform an analysis of the emotional content in more than 3,000 YouTube video transcriptions and comments associated with approximately 500 Amazon products. This study seeks to quantify these emotions and assess their statistical significance through statistical models such as logistic regression, providing insights into their relationship with Amazon ratings.

1.1 The Role of Emotions in Online Reviews

Emotions and sentiments play a crucial role in how people express opinions, and their importance in proving credibility and evaluating the helpfulness of online reviews are implied by various works[8][9]. Deriving sentiments from textual data is an exhausting process if done manu-

ally, but it can be automated using sentiment analysis techniques, including tools such as VADER[10] and SentiWordNet[11]. While sentiment analysis traditionally categorizes content into positive, negative, or neutral sentiments[11][12][13], these approaches oversimplify the complexity of human emotions and they should not be reduced to simple positive-negative distinctions[14]. Emotions are multi-faceted and provide a deeper layer of insight into consumer reactions to products. The consumers are heavily influenced by the emotions induced from events, marketing etc. [15]. For instance, an individual might feel excitement, frustration, or disappointment, emotions that go beyond the basic positive or negative sentiment.[16] Knowing the existence of this emotional complexity, drove this study to use more complex models to capture a better view of the emotions expressed in the comments and the videos. A dataset published by Google named GoEmotions[6] identifies 28 distinct emotions, allowing us to dissect the emotional content of YouTube reviews and comments with much finer granularity than traditional sentiment analysis. This RoBERTa based model[7] enables us to explore not just whether a review is positive or negative, but to understand the nuanced emotional states that reviewers experience when engaging with a product.

1.2 Research Objectives

The primary objective of this study is to determine if there is a correlation between the emotions expressed in YouTube reviews and comments and the product ratings on Amazon. Specifically, we aim to answer the following research questions:

1. Do the emotions expressed in the transcriptions of YouTube videos reflect the star ratings that the related products have on Amazon?

2. Do the emotions expressed in YouTube comments match the emo-

tions demonstrated in the transcriptions of the videos?

3. Can emotions be considered as relevant variables when measuring how well a product performs in the market?

By providing answers to these questions, this study enhances our understanding of how user-generated content can be used as an indicator of the quality of a product and the satisfaction of those who use it. The findings could potentially lay the groundwork for developing automated systems that evaluate the reliability of reviews based on their emotional content, thereby helping consumers make more informed purchasing decisions without relying solely on numerical ratings.

1.3 Methodological Approach

To explore these questions, this study utilizes a quantitative approach. First, a dataset comprising Amazon product URLs and their corresponding YouTube video reviews was manually curated. By leveraging YouTube's API[17], the English transcriptions of these videos and the associated comments were extracted. The videos span a range of product categories, excluding those that do not typically receive objective reviews, such as fashion and books.

The second stage involves the application of the fine-tuned RoBERTa model to classify the emotions in both the video transcriptions and the comments[7]. This model outputs a probability distribution across 28 emotions for each sentence in the review. These emotions are then mapped to six basic Ekman[16] emotion categories: anger, disgust, fear, happiness/joy, sadness, and surprise, along with a neutral category. The emotional content is aggregated at the video level, allowing us to compute the average intensity of each emotion for each video.

Next, a series of analyses were conducted to explore the relationship between these emotions and Amazon star ratings, including logistic re-

gression which was then applied to identify statistically significant emotions in relation to the Amazon ratings.

1.4 Significance of the Study

This thesis sheds light on the relationship between emotions and product ratings of two different platform. By moving beyond simple sentiment analysis and incorporating a broader spectrum of emotional categories, this research adds depth to our understanding of how online reviews influence consumer behavior. The findings have both theoretical and practical implications.

Theoretically, this study contributes to the growing body of literature on emotion recognition in natural language processing (NLP)[8][9][18]. By demonstrating the usefulness of emotion detection in analyzing product reviews, this research highlights the importance of moving beyond sentiment classification to a more nuanced understanding of consumer emotions.

In practice, the insights gained from this research could be applied to develop automated tools that identify emotional discrepancies in online reviews in order to assess their credibility. These tools could assist online retailers like Amazon detect possible instances of fabricated or manipulated reviews by flagging reviews that show emotions in contrast with the product's rating. The results may also guide the creation of systems that calculate product ratings relying on the emotional content of video reviews, thereby providing customers with a quicker option to evaluate items without having to read through lengthy text reviews or watch an entire video.

1.5 Structure of the Thesis

To provide a clear and organized exploration of this topic, the thesis is structured as follows:

- **Chapter 1: Introduction** – Provided an overview of the study's background, objectives, and significance, outlining the key research questions and the framework of the research process.
- **Chapter 2: Background** – Delves into the history and significance of Amazon and YouTube as platforms for e-commerce and product reviews, and reviews the existing literature on sentiment analysis, emotion detection, and consumer behavior.
- **Chapter 3: Methodology** – Details the data collection process, including the creation of the YouTube and Amazon dataset, preprocessing steps, and the emotion detection techniques used in this study.
- **Chapter 4: Experiments** – Describes the experiments conducted to assess the correlations between emotions and Amazon ratings, including an analysis of the granularity of emotions and the binary classification of ratings.
- **Chapter 5: Discussion** – Interprets the results of the experiments, addressing the significance of the findings of the thesis.
- **Chapter 6: Conclusion** – Summarizes the research findings and highlights the theoretical and practical contributions of the study, discusses the challenges faced during the research while offering insights into the potential for future applications of emotion recognition in e-commerce.

Chapter 2

Background

Before the introduction of the Internet, consumers had only a few possibilities for deciding on which product to buy, they would either visit a physical store to examine the products in person and judge by the looks or speak with the seller for assistance. They could have also discussed their opinions with others they knew, but that would have been very limited.[19] Today, the circumstances have radically changed due to the widespread usage of the internet.

Globally, consumers may benefit from the Internet equally regardless of where they reside. Online sales enable customers to compare prices, shop in a more convenient location, choose from a greater selection of products than in-store, and obtain additional details about the product without ever leaving the comforts of their home. The Internet is thought to be the most popular source for information collecting and is becoming more and more integrated into the lives of consumers. As a result, customers are no longer constrained by the products sold in physical locations and are able to make better informed purchases.[20]

The existence of internet has led to the creation of the electronic word-of-mouth (eWOM) concept, which is a powerful and influential form of communication that has distinct characteristics compared to traditional word-of-mouth (WOM). eWOM is marked by its scalability, speed, per-

sistence, and accessibility, which allow information to spread quickly and remain available indefinitely across online platforms.[21] However, eWOM poses unique challenges, such as the difficulty in assessing the credibility of anonymous sources, which can impact how consumers trust and adopt the information.

It has been shown by Hannan et al. that through the use of Natural language processing techniques, it is possible to efficiently identify features, determine the polarity of opinions, and provide structured summaries from large amounts of unstructured textual data, such as customer reviews and any other form of eWOMs.[22] These summaries are valuable not only to shoppers looking for quick insights but also to product manufacturers who need to understand customer feedback. Following the same pattern, Yin et al.[8] and Felbermayr[9] have successfully tried to categorize helpful eWOMs in the form of reviews which can affect a potential buyer using emotion recognition technqiues and signify the importance of the emotions in eWOM analysis.

To set the stage for the analysis and discussion that will follow, this chapter provides essential background information on the primary platforms involved in this study: Amazon and YouTube, providing a concise overview of each platform's role in the e-commerce ecosystem. The section 2.1 focuses on Amazon, detailing its evolution as a shopping platform and the importance of its user-provided reviews. The section 2.2 examines YouTube's role as a platform for video content, particularly focusing on its growing importance as a space for product reviews and consumer feedback. This is followed by a literature review which synthesizes key academic contributions and research findings related to the themes of this thesis. This chapter discusses existing studies on online reviews, sentiment analysis, the influence of emotions on consumer behavior, and the application of language models like BERT in analyzing textual data. Through this review, the chapter establishes the theoretical and empirical foundations that inform the subsequent analysis and experiments.

2.1 Amazon

As the internet began to transform commerce in the mid-1990s, specifically during the dot-com bubble, a number of companies emerged with the vision of revolutionizing how people shop,[23] with one of the most notable of these pioneers being Amazon.com. In July 1995, Amazon.com launched as a US internet bookseller. In October 1998, the firm established its first foreign websites, Amazon.de and Amazon.co.uk (Germany and United Kingdom, respectively), as it started to expand into other geographical regions. Notably, this regional growth was accomplished by the acquisition of two internet booksellers: The german plarform "Telebook", and "Book pages" situated in the United Kingdom. the goal from the very start was to make Amazon the largest online retailer of mass products.[24]

Figure 2.1: Amazon's website in 1995, the first year it was started.

In 1998, the company started selling toys, electronics, music, and videos as part of its ongoing product range expansion. In November

2000, following its initial transition from an online book seller to an online portal for customers to purchase goods, Amazon made the decision to expand its business plan to include a third-party marketplace.[25] This was an important moment in Amazon's history because the third-party marketplace would eventually become a major component of the business. As a result, the retail and the marketplace are now coexisting in the same space, which at the time was a really intriguing and novel idea. Put another way, customers could now decide whether to purchase directly from Amazon (the retail business model that had been in place up until this point) or from independent merchants. This means that Amazon began competing directly against third-party sellers inside the Amazon marketplace. Following this, Amazon started a very significant geographical and product line expansion. It is essential to point out that Amazon has grown to become one of the largest and most powerful businesses in the world as a result of its business model expansion and diversification. Customers may now purchase an extensive range of products on the platform with simple access and ease thanks to the third-party marketplace in particular. The marketplace's growth and financial success have been further aided by Amazon's "Fulfillment by Amazon" program launched in 2006, which offers fulfilment services to independent merchants. Smaller companies may simply offer their goods on the marketplace without worrying about the complicated processes of shipping and handling thanks to this.[26] It should be mentioned that a retailer has the option to offer their products across numerous marketplaces because of the Amazon Global Selling initiative. Through this program, retailers can sell in eighteen international marketplaces across five continents: North America (the US, Canada, and Mexico); Europe (the Netherlands, Spain, France, Italy, Germany, Sweden); the Middle East (the UAE, Saudi Arabia) and Asia-Pacific (Japan, Singapore, and Australia) and North Africa.[27] These days, small businesses have access to millions of customers who can order and get their goods in more than 100 countries. In fact, 300 million peo-

ple are thought to have shopped at Amazon locations globally (Quaker, 2022). Through its Amazon Web Services, Amazon Advertising, and Amazon Prime Video divisions, respectively, Amazon has branched out from its e-commerce business into other markets including cloud computing, advertising, and entertainment. Generally, Amazon's business model has grown and evolved steadily throughout its existence, constantly adjusting to new market opportunities and trends. Amazon allows each customer to leave a review for the products they purchase. Reviews play a vital role in Amazon sales, a study examined how online reviews shape sales of remanufactured phones, analyzing 50 thousands of reviews with NLP tools. The analysis suggested that Lengthier reviews and favorable sentiments were linked to higher sales, while usability, service quality, and cost-effectiveness were also found to affect purchasing decisions. Positive emotions like contentment and surprise appeared to drive sales upward, while anger had the opposite impact. The research uncovers key patterns in review characteristics that could be useful for enhancing sales strategies.[28] Another study researched the influence of online product reviews on consumer behavior, particularly in the context of e-commerce and focusing on Amazon as one of the platforms they obtained their data from. By analyzing how reviews, ratings, and their structures impact purchasing decisions, the research highlighted the crucial role reviews play in shaping trust and guiding consumers through the selection process. It was discovered that while reviews are valuable, the comparison between positive and negative reviews can be time-consuming due to their unstructured format.[29] Overall, Amazon is one of the most significant participants in the e-commerce retail sector. More importantly, it has been estimated that it will surpass every competitor globally by 2027. Amazon has evolved into a preferred resource for discovering new products by consumers. [30].

2.2 YouTube and Reviews

On April 23, 2005, Jawed Karim, Steve Chen, and Chad Hurley—three former PayPal employees—founded the video-sharing website YouTube. The first ever video, titled "Me at the Zoo"[31] was published to the YouTube on the same day by Jawed Karim. The goal of YouTube is to enable anyone to upload, watch, and share videos while making them accessible to everyone. There are many different types of content available in YouTube, such as music videos, TV snippets, tutorials, and video blogs. Following its launch, users grew from 30,000 per month to 12 million in a year. It saw on of the fastest growth rate in Internet history in 2006[32], which led to Google purchasing the business in October of that year for a huge sum of 1.65 billion dollars[33].

In 2010, the platform had more than two billion views every day. As a result, it became the world's first website focused on distributing videos online and the second search engine, behind Google. Meanwhile it was reported that YouTube serves twenty million visitors per month according to Nielsen/NetRatings.[34]

The ability for anyone to create a profile on YouTube, known as a channel, is what draws content creators to the platform. Owners of channels (whether people or organizations) can post a variety of contents, such as videos expressing their personal thoughts, reviews regarding specific businesses, products, or services. Additionally, users have the option to subscribe to other channels that interest them.[35] YouTube mandates that all users create a personal account in order to be followed by other individuals on the social network, even though uploading content is not required.

There are fifteen categories of content, according to the website: cars and vehicles, comedy, education, entertainment, film and animation, gaming, how to and style, music, news and politics, nonprofits and activism, people and blogs, pets and animals, science and technology,

sports, and travel and events. Thanks to this broad range of categories, users can discover the content they are looking for more easily. For example cosmetic and beauty influencers and reviewers typically fall into one of the following three categories: "how to and style", "people and blogs" or "entertainment."[36]

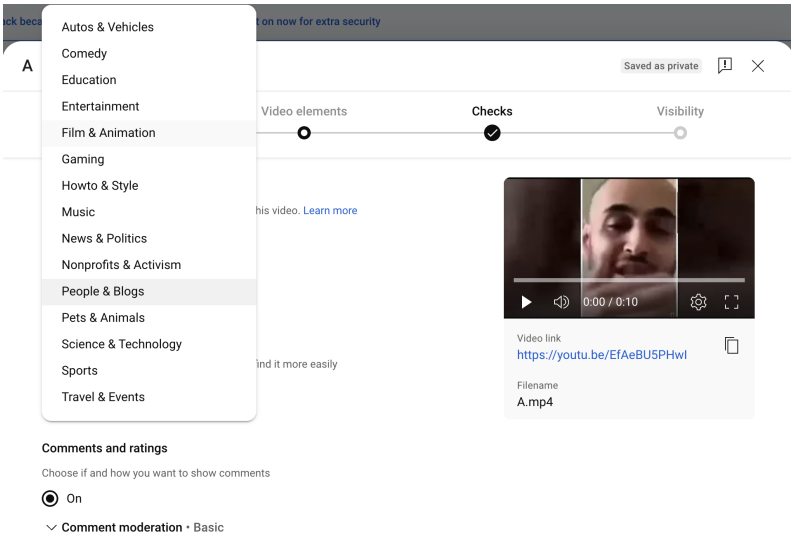

Figure 2.2: The categories of youtube as seen on YouTube studio.

2.2.1 "YouTubers" and their influence

According to YouTube press, every minute users upload more than 500 hours of videos[37] across plenty of channels. Naturally, not all of them have the same impact on their target audience; some are more popular than others. Assessing a channel's influence is rather straightforward: the most popular ones have more subscribers, views, and likes than the rest of the channels. We can ascertain who is considered most significant and influential in the context of an online community thanks to this data. When we discuss influence, we imply having the ability to engage the audience and ensure that they have devotion to a specific YouTuber. In this sense, the public's backing allows the YouTuber to establish a profitable business venture.

Countless video creators on YouTube with millions of followers have built solid relationships with the audience. Businesses are now able to leverage a mutually beneficial cooperation to convey marketing messages to their dedicated audiences. Influencer marketing, according to Chaffey & Ellis-Chadwick [38], is a type of online advertising that involves paying one or more YouTube influencers to collaborate on behalf of a brand in order to promote the latter's products and services to consumers of target. Influencer marketing can also be considered as an earned media channel, meaning that it aims to raise brand awareness by approaching the public through recognizable individuals

YouTube has considerable influence on how young consumers behave while making purchase decisions, particularly when linked with the personalities and endorsements of YouTubers. A YouTuber's fame, skill, credibility, and entertainment value are key factors which impact how viewers judge and purchase products. Product reviews rank as the most influential element among these, followed by the frequency of endorsements and engagement metrics like views, comments, and likes. Furthermore, a considerable correlation has been observed between the personalities of YouTubers and the efficacy of their product endorsements, highlighting the crucial role that YouTube plays in influencing consumer decisions.[39]

As a result of the growth of social media, a unique kind of celebrity known as "micro-celebrities" or social media celebrities has emerged.[40] These people, called Instagrammers and YouTubers, develop their online personas by actively interacting with their followers on social media [41]

Until recently, businesses were inclined to choose celebrities who had large social media followings, such as Cristiano Ronaldo, to endorse their products and services because of their capacity to reach and influence a sizable audience. But now that social media has taken off, anyone can post a message online, and it can circulate throughout the network and reach people worldwide. Therefore, businesses are increasingly using

social media influencers to promote and endorse their brands in addition to keeping working with traditional celebrities like actors, supermodels, and athletes to increase their brand awareness on social media.[42]

Penttinen et al. highlight the growing significance of consumer to consumer (C2C) video reviews—especially on platforms like YouTube—in influencing consumer purchasing choices. Although the study has mostly focused on textual reviews, it emphasises that video reviews are becoming increasingly important due to their capacity to foster parasocial interactions, or one-sided relationships between reviewers and viewers. According to the study, self-disclosure and interaction in video reviews can strengthen these parasocial connections, boosting the reviewer's reputation and positively impacting the purchase decisions of consumers. Notably, consumers who have less confidence in their decisions are more affected by these impacts. The results suggest that the emotional connection that video reviews establish has a major impact on consumer behaviour.[43]

Kim has shown how travel video blogs, or travel vlogs, on YouTube have had an immense effect on how consumers behave in tourism sector. These vlogs are quite interesting and offer immersive travel experiences, which influence the choices of destination and electronic word-of-mouth (e-WOM). The study demonstrates how motives like information seeking and entertainment drive the consumption of travel vlogs, resulting in increased engagement, through applying the use and gratification (U&G) perspective. Leveraging structural equation modeling and data from 300 respondents who had seen travel vlogs before travelling themselves, the study demonstrates that emotional involvement and perceived presence in these vlogs had a significant impact on travel intentions and e-WOM sharing. This research provides interesting facts regarding how new technologies, such as travel vlogs, influence consumer behavior[44]

Weiss claims that the potential for anyone to be an influencer nowadays is largely due to the power of word-of-mouth on the Internet. It

is important to keep in mind that consumers prefer to listen to credible sources—ordinary people—instead of corporations' endless flood of advertisements.[45]

YouTube is one of the most relevant sites to examine while studying 'User Generated Content', which denominate any videos that YouTube users upload to their channels[46] and every day, millions of videos are posted, covering a vast range of subjects, with YouTube announcing that their total watch time has surpassed 1 billion hours back in 2017[47]. This is one of the reasons why this thesis will be working on data from YouTubers' product reviews.

2.3 Emotions Classification and Analysis

The study of emotions and sentiments provides insightful data on the underlying perceptions, attitudes, and feelings expressed through text. Sentiment analysis has gained traction as an analytical tool across varied platforms[8][9], ranging from social media interactions to online product reviews. While NLP initially focused on structured textual sources, the advent of user generated content on platforms like YouTube and Amazon has expanded the scope of such analyses. By dissecting the sentiments and emotions expressed in comments and reviews, we aim not only to assess consumer feedback but also to explore deeper social and behavioral patterns. The analysis of emotions require a framework of recognizing different emotions. The classification of emotions has evolved over time, with each new classification models trying to represent more complex emotions. While modern methods emphasise dimensional or continuous models that take into account complex emotional states, early models were mostly based on psychological theory and concentrated on a small number of basic emotions.

An influential emotion classification model in the field is Paul Ekman's basic emotions theory[**?**], which proposes six universal emotions: hap-

piness/Joy, sadness, anger, fear, disgust, and surprise. This model has been criticised for oversimplifying human emotional experiences despite its widespread acceptance because many emotions go beyond the suggested categories.

Mehrabian and Russell[**?**] introduced The PAD (Pleasure-Arousal-Dominance) model, which provided a 3 dimensional approach to emotion classification. The three continuous axes which are utilised to classify emotions consist of:

- **Pleasure** indicating the degree of positivity or negativity of the emotion;
- **Arousal** indicating the level of activation or energy associated with the emotion;
- **Dominance** indicating the extent to which one feels in control or overpowered by the emotion.

In comparison to categorical models, this framework provides a more flexible representation of emotions, demonstrating also the intesity of the emotions..

Plutchik proposed the Wheel of Emotions based on the previous frameworks. It organises emotions into eight basic bipolar pairs: joy-sadness, trust-disgust, fear-anger, and surprise-anticipation[**?**]. This model suggests that emotions are not isolated entities but exist in relation to one another, differing through combinations and intensities. Plutchik's model introduced the idea that complex emotions are mixtures of basic ones, which has influenced later models that aim to capture a wider emotional spectrum in textual analysis.

Over time, more nuanced classifications emerged, recognizing the limitations of basic emotion categories. The Geneva Emotion Wheel (GEW)[**?**] reflects a greater range of emotions, mapped onto two axes: control and pleasantness, which enhances upon the earlier classification models. This

representational model makes it possible to represent emotions in a more dynamic way.

In computational sentiment analysis, categorical and dimensional models often intersect, with tools like DeepMoji and GoEmotions leveraging both approaches to recognize emotions from textual data. The GoEmotions dataset is one example which covers 27 emotion categories across a wide range of intensities and valences, expanding the possibilities of fine-grained emotion detection in user-generated content[6].

As emotion classification models continue to evolve, they increasingly reflect the complexity of human emotions, incorporating both traditional psychological insights and modern computational techniques.

Having a background for the categories of emotions allows for creation of tools which classify different emotions. SentiWordNet[11] is a tool developed to create a tool to help detect whether words in a text express positive, negative, or neutral sentiments, built on WordNet[48], which is a large lexical database of English words. Instead of evaluating just the word itself, the authors assessed each synset which themselves are a group of words or phrases in WordNet that share the same meaning. For each synset, SentiWordNet assigns three scores, positivity, negativity, and objectivity, indicating how likely the word is to convey a positive, negative, or neutral sentiment. This tool is still one of the widely used tools for any sentiment analysis task[49][50].

SentiStrength[12] is also a similiar attempt like SentiWordNet, it is an algorithm that uses a combination of a sentiment word list, machine learning techniques, and methods to handle non-standard spellings and abbreviations commonly found in online platforms like MySpace. The system achieved significant accuracy, predicting positive emotion with 60.6% accuracy and negative emotion with 72.8%, making it a useful tool for analyzing sentiment in any use cases.

To examine another method of sentiment classification, Hazarika et al.[13] explore how TextBlob, a natural language processing library, can

be applied to analyze the sentiment of tweets from Twitter. Their goal was to classify the polarity of tweets (positive, negative, or neutral) by applying TextBlob's sentiment analysis capability, which relies on a pretrained NaiveBayes classifier.

A simple yet effective model for analyzing sentiment in textual data is VADER[10] (Valence Aware Dictionary for Sentiment Reasoning), a rule-based tool specifically designed for the short, informal nature of social media text. The authors compared VADER's performance against other widely used sentiment analysis tools and found that it outperformed both human raters and more complex models in various contexts, achieving high accuracy in classifying the sentiment of tweets and other contexts.

To improve sentiment, emotion, and sarcasm detection by leveraging the vast amount of emojis used in social media posts as noisy labels for distant supervision, Felbo et al. introduced DeepMoji[51], a model pretrained on 1.2 billion tweets containing emojis, which significantly enhances performance across various NLP tasks. Their model outperformed existing benchmarks in emotion and sarcasm detection at the time, demonstrating the effectiveness of emoji based supervision in generating rich, transferable text representations

Siersdorfer et al. examined YouTube comments and predict their usefulness based on community feedback and sentiment.[52] Using a dataset of over 6 million comments from 67,000 videos, they examined the relationship between comment sentiment, community ratings, and video topics. By leveraging the SentiWordNet[11] and machine learning classifiers, they successfully demonstrated that it is possible to predict the community acceptance of comments based on their content and sentiment.

Nandwani et al.[53] attempted to give an extensive overview of the main techniques used for sentiment and emotion analysis. They investigate the models of emotion detection, each level of sentiment analysis (document, sentence, and aspect levels), and the processes involved

(preprocessing, feature extraction, analysis techniques, etc.). The study effectively brings attention to the difficulties faced by researchers in these fields, such as managing unstructured data, evaluating linguistic ambiguities, and identifying several emotions in a single sentence.

The SenTube dataset is an attempt at analyzing user-generated comments on YouTube videos, annotated for both information content and sentiment polarity. With the use of this dataset, classifiers for key natural language processing (NLP) applications like sentiment analysis, text classification, spam identification, and informativeness prediction can be developed. SenTube is a testament to the feasibility of using YouTube's videos and comments as a source of information.[54]

Shah and Parekh analyzed user sentiments expressed in YouTube comments regarding various productivity tools.[55] They employed the VADER sentiment analysis tool along with machine learning algorithms like Naive Bayes and Random Forest to classify sentiments as positive, neutral, or negative. The study achieved notable accuracy rates of 84.01% with Naive Bayes and 91.34% with Random Forest, demonstrating the latter's superior performance in accurately identifying sentiments, particularly for neutral and negative sentiments

Sahu et al. suggest an innovative approach for recommending future films by combining a hybrid recommendation system with sentiment analysis. The research fills the gap in conventional recommendation systems, which usually concentrate on already-released films, by concentrating on unreleased films that are accessible through trailers. The study initially obtains the general sentiment and expected rating of these unreleased films by applying the VADER sentiment analyzer on YouTube comments on Netflix movie trailers. Subsequently, it incorporates these data into a hybrid recommendation system, proposing upcoming films based on user preferences. [56]

ADIGÜZEL focuses on how user interaction with YouTube game reviews affects the retail sales of video games. The researchers analyzed 140

YouTube reviews of well-known games that were published between 2010 and 2019 in an attempt to determine how social media metrics—like/dislike ratios, sentiment, and total views—affect sales. According to the study, YouTube reviews have a considerable impact on sales, but not as much as traditional consumer and critic text reviews, with customer reviews having the biggest positive influence. It is interesting to note that sales were not considerably impacted by the tone of YouTube reviews or whether they were written for companies or by individuals. These results demonstrate how important it is for both text and video reviews to influence consumer behavior and forecast sales in the video game market.[57]

Pradhan concentrate on evaluating public opinion regarding video content by utilizing Vader tool to analyze comments' sentiment on YouTube. This analysis assists content creators in better understanding user feedback and making improvements to their videos by categorizing comments into three groups: positive, negative, and neutral. Overall, the findings show that YouTube comments are a useful instrument for evaluating audience response and show that, probably as a result of the often entertaining nature of YouTube content, positive comments frequently exceed negative ones. [58]

Ray et al. combined sentiment and emotional analysis in an attempt to boost the ability of prediction of ratings using only social media comments. Social media lacks distinct rating indicators, which makes it difficult to evaluate user comments, in contrast to traditional review sites where ratings are clearly provided. Through an analysis of 3,509 tweets about e-learning, travel agencies, and online food delivery, the study was able to achieve significant results in prediction accuracy. R libraries such as "sentimentr" and "syuzhet" made it possible to analyze different emotions in detail, which improved the ability to understand user feedback and create recommender systems that perform better. [59]

There are multiple attempts done at classifying the sentiment of Amazon product reviews, one example is the work of Hawlader et al.[60]

which is aimed to evaluate the accuracy of different supervised learning algorithms for the task. The researchers explored a variety of feature extraction methods consisting of Word2Vec[61], TF-IDF, and Bag of Words, together with classifiers including Naive Bayes, Support Vector Machines, Decision Trees, Random Forest, Logistic Regression, and Multi-Layer Perceptron (MLP) and according to their examinations the MLP classifier and the Bag of Words model together produced the best accuracy.

Wassan et al. evaluated the sentiment of Amazon product reviews using machine learning approaches by which, they analyzed 28000 user reviews across 60 different product categories.[62] The sentiment analysis classified reviews as positive, neutral, or negative based on polarity and subjectivity measures. Their findings highlighted that a large proportion of the product reviews showed a positive sentiment. They showed that this method of interpretation of reviews can be a helpful tool for marketers to understand consumer behavior and preferences, which can aid in improving product quality and customer satisfaction.

In an attempt to classify amazon reviews for a specific product based on the sentiment, Kausar et al. examined the subject [63] utilizing two models, Decision Trees and Logistic Regression, and concentrated on reviews of a watches brand named "Titan Men Watches", obtained from amazon.in. The Bag of Words method was implemented to process the textual reviews, and both models performed quite well. The Logistic Regression model had an accuracy of 94%, whereas the Decision Tree model had an astonishing 99% accuracy. These results demonstrate how machine learning techniques may be used to efficiently categorize customer sentiments from product reviews.

Chaehan So investigated the relationship between ratings and emotions in Amazon product reviews, as well as any bias in the machine learning models' interpretation of these emotions. According to the study, the most significant class of emotions in predicting product ratings were "joy"

and negative emotions. In spite of their overall good performance, further investigation indicated that the models may not have been entirely fair. The models tended to be biased towards positive ratings. For instance, the model tended to predict higher ratings even when the emotion "joy" was less frequent. This implies that skewed results may have resulted from the model's overemphasis on positive emotions and underemphasis on negative ones.[64]

Qorich and El Ouazzani explore the analysis and classification of sentiments in Amazon reviews of products as either positive or negative using a Convolutional Neural Network (CNN) model. Through the use of several word embedding techniques, the authors evaluate their CNN model in an effort to increase the accuracy of sentiment analysis. The CNN model presented in the paper outperformed deep learning and conventional machine learning techniques, achieving 90% accuracy. The study suggests that training word embeddings particularly for large-scale datasets such as Amazon reviews can lead to improved performance compared to utilizing pre-trained embeddings. [65]

NLP models have significantly improved with advances in machine learning and deep learning. They work by modifying textual material to facilitate additional analysis. As the AI field has gone further and processing capacity have increased, NLP models have grown in popularity and complexity. These models have evolved over time from simple statistical methods to more complex and computationally intensive models[66]. The introduction of the Transformer architecture was a recent development that altered the field of natural language processing[1]. The Transformer architecture was at first created for the machine translation task for which it performed astronomically better than previous methods like CNN or Deep Learning based models. following the transformers' introduction, a new model called BERT[4] (Bidirectional Encoder Representations from Transformers), based on the same architecture, was developed. BERT is a pretrained model that learns contextual

relationships between words in a sentence by training on large amounts of text data. Upon its introduction, it achieved excellent performance on various sentence-level and token-level tasks. Building on the concept of pretrained Transformers, OpenAI created the GPT (Generative Pretrained Transformers) models[67]. Unlike BERT, which uses only an encoder, GPT based models are capable of generation.

Liu et al. introduce a new, BERT inspired model named RoBERTa which outputs even greater performance by adjusting the BERT training process. The researchers discovered that by modifying important variables like training duration and data size, BERT could be greatly enhanced. By utilizing longer text sequences and eliminating some training objectives, RoBERTa outperforms BERT and achieves better performance on multiple key benchmarks. These enhancements demonstrate how carefully modifying training techniques can increase a model's performance, putting RoBERTa above BERT and establishing a new benchmark in the industry.[3]

Üveges et al. and Xiangyu et al. have focused on developing a transformer-based model for analyzing emotions and sentiments by fine-tuning a BERT model, with HunEmBERT[68] created for Classifying Sentiment and Emotion in Political Communication and BERT-ERC[69] for the task of emotion recognition in conversation. Both of the studies demonstrate how the resulting models highlight shortcomings but also shows how large-scale language models can be fine tuned successfully for certain applications.

To analyze Amazon product reviews, Ali et al.[70] examined the task employing a wide range of machine learning, deep learning, and transformer based models. They focused on simple models such as Logistic Regression, Random Forest, and continued with more complex models like CNNs, Bi-LSTM, BERT, and XLNet. The results of their comprehensive analysis revealed that the BERT model outperformed others, highlighting the potential and effectiveness of transformer based models

for sentiment analysis tasks.

Demszky et al. present GoEmotions, an extensive dataset of 58,000 English comments on Reddit that has been carefully classified as Neutral or one of 27 emotion categories. The goal of this dataset is to improve the comprehension and linguistic expression of emotions, which is essential for creating humane chatbots and identifying inappropriate online behavior. The study demonstrates that the dataset, which was validated using Principal Component Analysis, offers accurate emotion labels and functions well in a variety of contexts and emotion frameworks. With a BERT-based model, the study's average F1-score was 0.46, suggesting a great deal of room for improvement in the future. Overall, GoEmotions establishes a solid basis for emotion prediction while emphasizing areas for improvement and extension into other languages and cultural contexts. [6]

In this chapter, we explored how NLP tools are used for sentiment analysis, particularly in the context of user generated content on platforms like Amazon and YouTube. These tools and methods help break down great amounts of unstructured texts such as customer reviews or video comments, into insights about the products and trends. Models like SentiWordNet, VADER, and newer transformer based models such as BERT have shown how emotions and opinions in text can be captured, which is essential for understanding consumer thought process. For example, they help us identify trends in reviews, measure customer satisfaction, and understand how product features are perceived.

However, the mentioned methods do have limitations as many sentiment analysis tools still struggle with nuances, such as sarcasm or ambiguous language, which can skew results. Additionally, the accuracy of these models can be affected by the context in which words are used, especially when dealing with slang, mixed emotions, or complex sentences. Pretrained models, while effective, often require fine-tuning to adapt to specific domains or languages, making them less universally reliable out

of the box, but are the best tools available today.

To conclude this chapter, we have walked through the significant roles that Amazon and YouTube play in the e-commerce ecosystem, especially regarding product reviews and consumer feedback. We also reviewed key studies on online reviews, sentiment analysis, and the emotions behind consumer decisions. By understanding both the potential and limitations of sentiment analysis and emotion detection, we have laid the foundation for the analysis and experiments that follow in the subsequent chapters.

Chapter 3

Methodology

In this chapter we delve into the process of gathering our data and describe our methods of research. First in section 3.1 we demonstrate our raw data gathering process and our criteria for adding a product, then in section 3.1.1 it is shown how the actual data is extracted from the captured list of links from previous section, showing how from the links we construct our dataset. In section 3.1.2 the dataset and each field is explained in detail. In the end we go into details over the preprocessing component in section 3.2 and provide insights on how each subprocess help with refining the data and in section 3.3 we explain how we augmented our dataset with emotion data and describe the tools used.

3.1 Data Collection

To gather the dataset, product links from Amazon.com were manually collected. The first criterion for product selection was category: we focused on categories where products are commonly reviewed on YouTube and tend to receive more objective reviews. As a result, categories like "Fashion", "Books" and "Digital Content" were excluded. After determining the categories, we aimed to select products randomly. However, Amazon's ranking system complicates true randomness, as it promotes

'featured' products more prominently. To address this, we utilized an unmentioned method found on reddit[71] that forces Amazon to sort products by average review ratings, allowing us to select products at random while ensuring a wider range of ratings. A sample of a product's amazon page is shown in figure 3.1

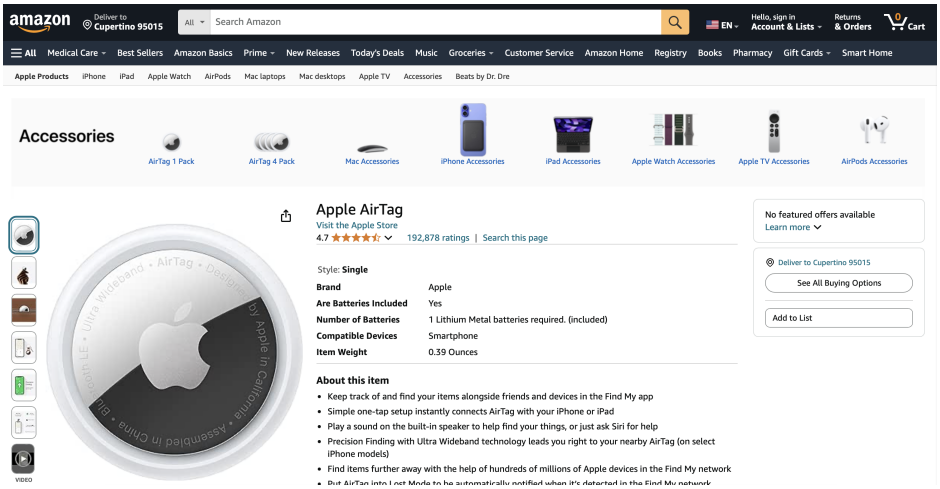

Figure 3.1: The Amazon page for the product 'Apple AirTag'.

For each randomly selected product, the product title was searched with the word "review" appended to locate relevant YouTube reviews as shown like the example in figure 3.2. Up to the first 10 videos recommended by YouTube were selected. Products without any available reviews were discarded from the dataset.

To prevent personal data from influencing the search results, a VPN and a web browser with a cleared history were used to conduct the searches and retrieve results. With the product links gathered in an Excel file, the next step was to extract the relevant data from the links. It is necessary to determine which specific data points would be collected from both Amazon and YouTube.

3.1.1 Data Selection

Amazon: The following data were selected for extraction:

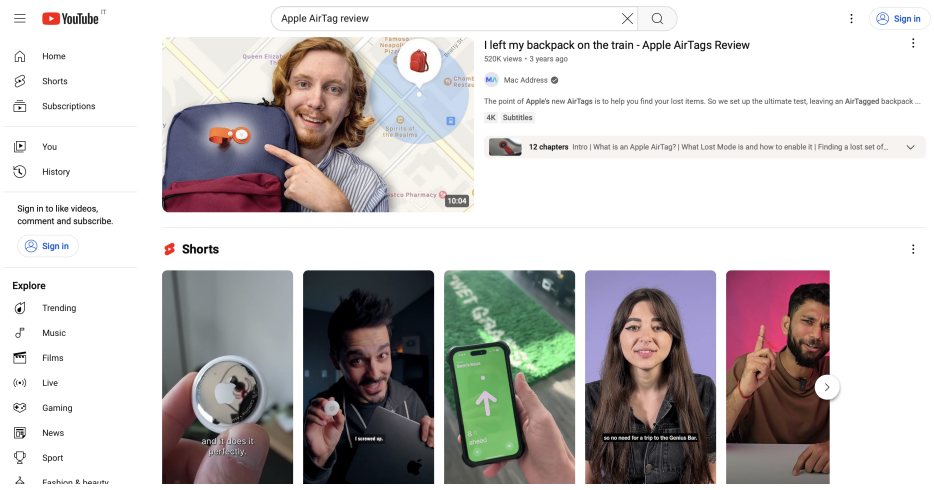

Figure 3.2: An example of a search for reviews for the product 'Apple AirTag'.

- Average rating of the product (on a scale from 0 to 5)
- Distribution of review counts based on the number of stars (from 0% to 100%)
- Product title (text)
- Total number of ratings
- Product category

YouTube: The primary data collected include video comments and transcripts, along with additional metadata as follows:

- Subscriber count of the posting channel
- Number of likes on the video
- Video duration
- Total number of comments

To collect this information, a Python 3 script was written which can be seen below. The "BeautifulSoup[72]" library was employed to simplify the process of scraping data from Amazon's web pages.

For retrieving metadata such as the number of likes, comments, replies, and other relevant information from YouTube, the "python-youtube"[73] library was utilized, which operates on the YouTube Data API. Additionally, "Pandas"[74] was used for data manipulation, visualization, and reading CSV files. Another valuable tool, "YouTube-transcript-API"[75], was employed to extract video transcripts from YouTube.

Since the YouTube-related libraries require an API key, one was generated using the Google Developer Console[76] to access YouTube data. To ensure organization and maintain structured data, the results were saved in a JSON file. The structure of this JSON file is illustrated in the figure 3.3.

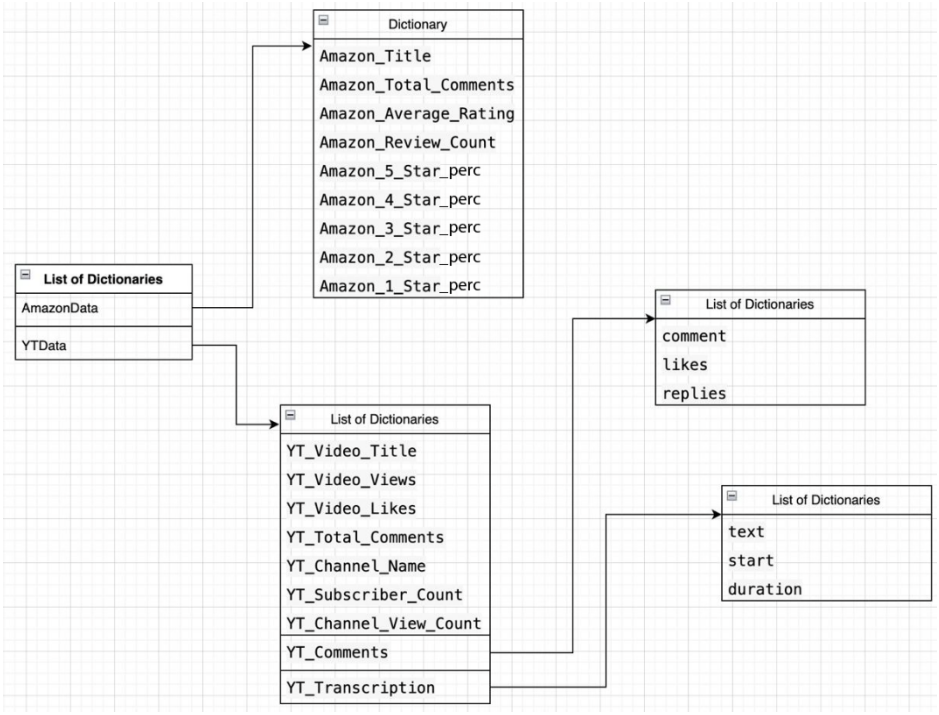

Figure 3.3: A diagram depicting the structure of our dataset.

At the end, A total of 498 products, 3,451 videos, and 137,248 comments were collected. Due to the large volume of data and the fact that each product is associated with multiple videos, it is not feasible to display all related videos here. Instead, a reduced example is provided in

the table 3.1.1, showcasing the first product, its first video, and the first comment for demonstration purposes.

Amazon Title	Apple AirTag
Amazon Total Comments	170,165 ratings
Amazon Average Rating	4.7 out of 5 stars
Amazon Review Count	170,165 ratings
Amazon 5 Star percentage	84%
Amazon 4 Star percentage	8%
Amazon 3 Star percentage	3%
Amazon 2 Star percentage	1%
Amazon 1 Star percentage	3%
YT Video Title	I left my backpack on ...
YT Video Views	510,633
YT Video Likes	24,246
YT Comments	
Comment	No one at the airport: "I can ...
Likes	2,196
Replies	22
YT Total Comments	1,426
YT Transcription	Your attention please.
YT Channel Name	Mac Address
YT Subscriber Count	610,000
YT Channel View Count	122,461,013

Table 3.1: Amazon and YouTube Data example.

3.1.2 Dataset Information

Our dataset consists of two main Lists: **AmazonData** and **YouTube-Data**, each containing dictionaries with their relevant fields.

- **AmazonData:** List of Dictionaries, each element represents one product and for each product it has:
  - **Amazon_Title**: The title of the product as listed on the Amazon page.
  - **Amazon_Average_Rating**: The average rating of the product on a scale from 0 to 5, as displayed on the Amazon page.

- **Amazon_Review_Count**: The total number of reviews for the product.
- **Amazon_5_Star_perc**: The percentage of 5-star reviews.
- **Amazon_4_Star_perc**: The percentage of 4-star reviews.
- **Amazon_3_Star_perc**: The percentage of 3-star reviews.
- **Amazon_2_Star_perc**: The percentage of 2-star reviews.
- **Amazon_1_Star_perc**: The percentage of 1-star reviews.

- **YouTubeData:** List of Dictionaries with each element being a list of videos for a product, and for each video it has:
  - **YT_Video_Title**: The title of the YouTube video.
  - **YT_Video_Views**: The total number of views for the video.
  - **YT_Video_Likes**: The number of likes for the video.
  - **YT_Total_Comments**: The total number of comments on the video.
  - **YT_Channel_Name**: The name of the YouTube channel.
  - **YT_Subscriber_Count**: The subscriber count for the YouTube channel.
  - **YT_Channel_View_Count**: The total view count of the channel.
  - **YT_Comments**: A list of dictionaries where each dictionary contains:
    - **comment**: The text of the comment.
    - **likes**: The number of likes on the comment.
    - **replies**: The number of replies to the comment.
  - **YT_Transcription**: A list of dictionaries where each dictionary contains:
    - **text**: The transcribed text of the video.

- **start**: The starting timestamp of the text segment.
- **duration**: The duration of the text segment.

3.2 Preprocessing

Preprocessing is a crucial step in any NLP task, as it ensures that the input data is standardized, cleaned, and prepared for effective analysis. In this stage, the main goal is to reduce noise in the data by removing irrelevant information, normalizing text, and ensuring consistency across the dataset. Preprocessing is specifically needed here since we focus on cleaning text sourced from online conversations, where informal language, slang, and irrelevant tokens are prevalent. These textual elements, such as excessive punctuation, emojis, URLs, and other artifacts of online discourse, can obscure the emotional content and hinder accurate analysis.

3.2.1 Preprocessing Pipeline

To obtain a refined text, we designed a comprehensive preprocessing pipeline that addresses these challenges in a systematic way. The preprocessing pipeline consists of several key stages that incrementally transform the raw text into a cleaner, more uniform form. The full pipeline is illustrated in Figure 3.4 and each step is described in detail below

1. **Emoji Substitution**: Emojis are ubiquitous in online communication, functioning as visual symbols that convey a broad spectrum of emotions without the use of traditional words. They act as emotion-rich resources, enhancing the emotional tone of text in subtle but significant ways. In this step, we replace each emoji with its corresponding textual meaning. For instance, a smiley face emoji would be translated into the word "happy." This substitution allows the underlying emotion conveyed by the emoji to be

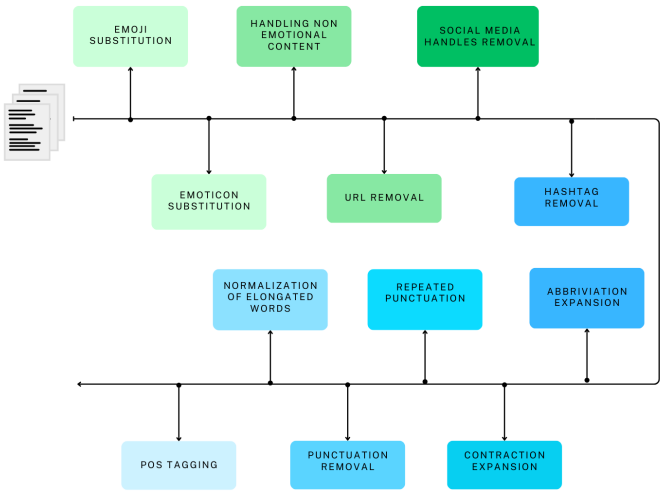

Figure 3.4: A visual representation of the preprocessing pipeline designed for this task.

preserved in a textual format, ensuring that the emotional nuances are captured and ready for further analysis.

2. **Emoticon Substitution**: Emoticons are character-based representations of facial expressions, such as :-) or :'(, commonly used in informal text communication. Like emojis, they hold emotional value and contribute to the sentiment of a sentence. Using a predefined dictionary of emoticon meanings, each emoticon is substituted with its textual equivalent. This ensures consistency and uniformity in the emotional content, allowing emoticons and emojis to be treated similarly during preprocessing.

3. **Removal of Parenthetical Content**: Parenthetical content, typically added by users in transcriptions and comments, often provides additional explanations or clarifications. However, this content usually lacks emotional relevance, serving instead to describe actions or provide context unrelated to sentiment. Regular expressions are employed to systematically detect and remove all text within parentheses. This step ensures that our focus remains solely

on the emotional aspects of the text, free from extra information.

4. **URL Replacement**: URLs (Uniform Resource Locators) frequently appear in online discourse, providing links to external websites or resources. While useful for information retrieval, URLs do not contribute to the emotional tone of a conversation. Therefore, all URLs are replaced with the placeholder term "website", maintaining a standardized format across the dataset without introducing irrelevant content. This substitution prevents URLs from interfering with sentiment analysis while keeping the text uniform.

5. **Mention and Social Media Handle Removal**: Social media interactions often include mentions or tags, denoted by the "@" symbol, to reference other users or accounts. These mentions serve a functional role in digital communication but do not contribute emotional information. Consequently, all instances of social media handles or mentions are entirely removed during this stage of preprocessing to streamline the text and focus exclusively on emotionally significant content.

6. **Hashtag Elimination**: Hashtags (preceded by the "#" symbol) are commonly used on platforms like YouTube and Twitter to categorize content or highlight topics. While they provide context, they do not express emotions directly and are therefore not relevant to sentiment detection. All hashtags are systematically removed from the text to eliminate noise and allow for a cleaner analysis of emotional content.

7. **Abbreviation Expansion**: Abbreviations and internet slang, such as "LOL" (Laughing Out Loud) or "BRB" (Be Right Back), are widespread in informal online communication and often serve as potent markers of emotion. To ensure that these indicators are captured accurately, each abbreviation is expanded to its full meaning

using a dictionary of internet slang. This expansion enhances the interpretability of the text by making explicit the emotional tone conveyed through abbreviations, thus preserving the richness of the sentiment data.

8. **Contraction Expansion**: In everyday communication, contractions such as "don't" (do not) or "can't" (cannot) are frequently used to shorten phrases. To create a uniform textual format, all contractions are expanded using a comprehensive dictionary of contracted forms. This transformation standardizes the text, ensuring that all words are in their full form, which can improve both the accuracy of sentiment analysis and the consistency of the data.

9. **Handling Repeated Punctuation**: In informal writing, repeated punctuation marks, such as "!!!" or "???" often indicate strong emotions like excitement, frustration, or confusion. These patterns provide subtle but important cues about the emotional intensity of a sentence. To retain this emotional emphasis, we replace repeated punctuation with the phrase "with strong feelings." This allows the text to reflect the heightened emotional state while avoiding excessive punctuation that may complicate further processing.

10. **Punctuation Removal**: Apart from repeated punctuation, which carries emotional weight, most punctuation marks—such as commas, periods, and colons—are not essential for emotion detection. Therefore, standard punctuation is removed to simplify the text, reducing its complexity and focusing solely on the emotional content. This step aids in streamlining the data for the subsequent stages of sentiment analysis.

11. **Elongated Word Normalization**: Elongated words, such as "cooooool" instead of "cool," are often used in social media to emphasize a particular sentiment or emotion. This form of empha-

sis is a crucial marker in sentiment analysis. In the preprocessing stage, we normalize elongated words by replacing them with their standard form prefixed by "very." For example, "coooool" becomes "very cool," thus accounting for the intensification while maintaining a standardized textual format.

12. **Part-of-Speech (POS) Tagging**: The final stage of preprocessing involves POS tagging, a linguistic annotation process that assigns syntactic labels (such as noun, verb, adjective) to each word in a sentence. Adjectives, in particular, are closely associated with emotional content, and their presence is often indicative of sentiment. Using tools like NLTK, Stanford-NLP, and spaCy, we POS tag the entire dataset, facilitating further filtering if needed. The POS tags are standardized into 17 groups, following the universal POS tagset. Cheng et al. introduced a method which combines the outputs from multiple POS tagging libraries and by utilizing it[77], we ensure high accuracy and compatibility in the tagging process, setting the stage for more nuanced emotional analysis in the subsequent steps.

3.3 Emotion Detection

Using Facebook's RoBERTa base model [2], fine-tuned on the GoEmotions dataset developed by Google Research [6], we tried to extract the emotion of each sentence in the videos and comments associated with each product. The GoEmotions dataset provides a granular classification of emotions, identifying 27 distinct emotions along with one neutral category, for a total of 28 emotions, as depicted in Figure 3.5. This fine grained emotion labeling expands the traditional six basic emotions: sadness, happiness, fear, anger, surprise, and disgust, and it does it mostly by dissecting the happiness emotion into finer emotions like joy, love, etc.

Given the granularity of the emotion classification in the GoEmotions

Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

Figure 3.5: Go emotions dataset emotions classifications.

dataset, our next step was to map the 28 emotions identified by the RoBERTa model into the more commonly accepted six basic emotions. According to the GoEmotions paper, these 28 emotions can be grouped into the six core emotions: anger, surprise, disgust, joy, fear, and sadness, along with a neutral category.[6] This mapping is shown in Table 3.2.

Basic Emotion	**GoEmotions Categories**
Anger	Anger, Annoyance, Disapproval
Surprise	Surprise, Confusion, Curiosity, Realization
Disgust	Disgust
Happiness/Joy	Love, Joy, Admiration, Relief, Approval, Pride, Optimism, Desire, Gratitude, Caring, Amusement, Excitement
Fear	Fear, Nervousness
Sadness	Sadness, Grief, Remorse, Embarrassment, Disappointment
Neutral	Neutral

Table 3.2: Emotion groupings of the 28 emotions from the GoEmotions dataset into six basic categories.

Once we established these groupings, the next step was to extract the emotions present in the sentences and comments from each video. The fine-tuned RoBERTa model was used to predict the emotions associated with each sentence, providing a list of probabilities that represent the proportions of each emotion expressed. Since these probabilities are distributed across all 28 emotions, we then consolidated them according to

the groupings shown in Table 3.2, deriving the six basic emotions and neutral.

One important decision during this process was the treatment of neutral sentences. Sentences classified as predominantly neutral do not contribute meaningfully to our analysis, which is focused on detecting emotional impact on product ratings. Therefore, we discarded sentences with neutral classifications to ensure our dataset reflects only meaningful emotional expressions.

For the remaining sentences, we aggregated the emotional proportions first by performing a sentence-level emotion extraction where for each sentence in a video or comment we derived a list of proportions representing how much each emotion is expressed. Then For each video, we summed the emotion proportions across all sentences, and then normalized these sums by dividing by the total number of sentences in the video. This gave us the average proportion of each emotion expressed within the video.

With these steps, we recreated our dataset to capture the emotional landscape of each product video. For every video, we calculated the average percentage of each of the six basic emotions (anger, surprise, disgust, joy, fear, and sadness) that were expressed along with the neutral category. Additionally, we paired this emotional profile with the corresponding product's Amazon rating, along with supplementary metadata such as the number of raters, video likes, and other relevant features.

By converting the 28 granular emotions into broader categories, we were able to maintain the richness of the emotion data while simplifying the analysis and drawing connections between emotional content and product success metrics such as ratings. This methodology also allows for the exploration of which emotional groups are most predictive of product ratings, which we will explore in subsequent sections.

Chapter 4

Analyses

In this chapter, we explore the experimental results from our study, offering a detailed interpretation and analysis. We will examine the data obtained in the previous chapter, assess its quality, identify correlations, detect multicollinearity, and apply various models to derive the results that will be discussed.

4.1 Descriptive Analysis

When analyzing our dataset, we observed some interesting patterns in the distribution of words, especially adjectives, which are key indicators of emotional content. From the transcriptions, there are 4,581,168 total tokens, and of those, 348,882 are adjectives, making up about 7.6% of all the words. As expected, most of the words used are fairly neutral and don't contribute much to emotion detection. For instance, the determiner 'the' was the most frequently used, with 215,289 occurrences. Interestingly, the most common adjective was 'little,' appearing 14,897 times. This suggests that words describing size or quantity are frequently used in product reviews, even though adjectives form a small portion of the overall text.

In contrast, when looking at the comments section, we have 3,576,551

tokens, and 346,773 of them are adjectives, accounting for about 9.7%. This higher percentage of adjectives in comments might indicate that people express more emotions in the comment section compared to the transcriptions from videos. Similar to the transcriptions, 'the' was still the most used word, but its frequency dropped to 163,818. This reduction might reflect the more casual writing style often seen in online social media, where slang or informal language is more common. Additionally, the most frequently used adjective in the comments was 'strong,' appearing 39,708 times. This is likely influenced by our preprocessing step, where repeated punctuation, used to emphasize emotions, was replaced with the phrase 'with strong feelings.' This suggests that people tend to express more intense emotions in comments compared to video transcriptions.

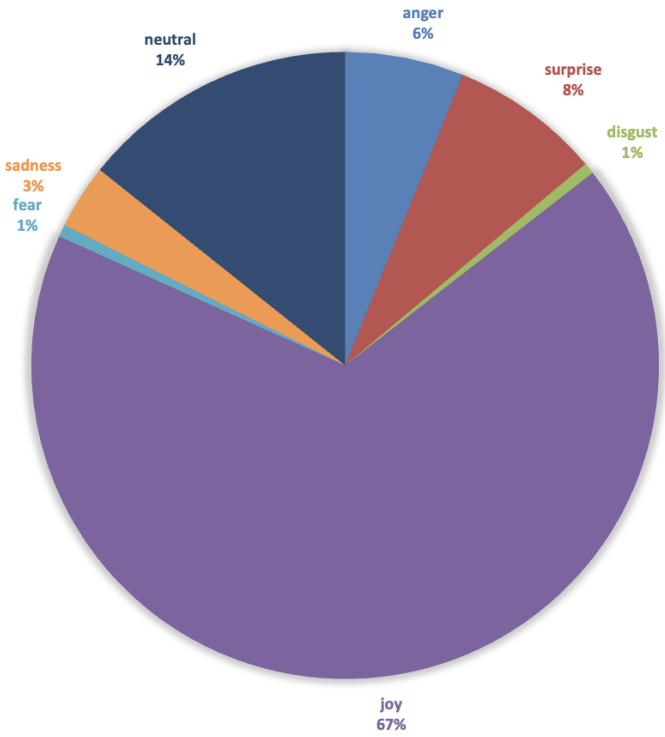

Figure 4.1: Pie chart showing the average expressed emotions in the videos.

From the emotion expressed inside the videos, as seen on the figure 4.1, most of them are positive, with joy taking 67% of average expressed

emotions in all of the videos.

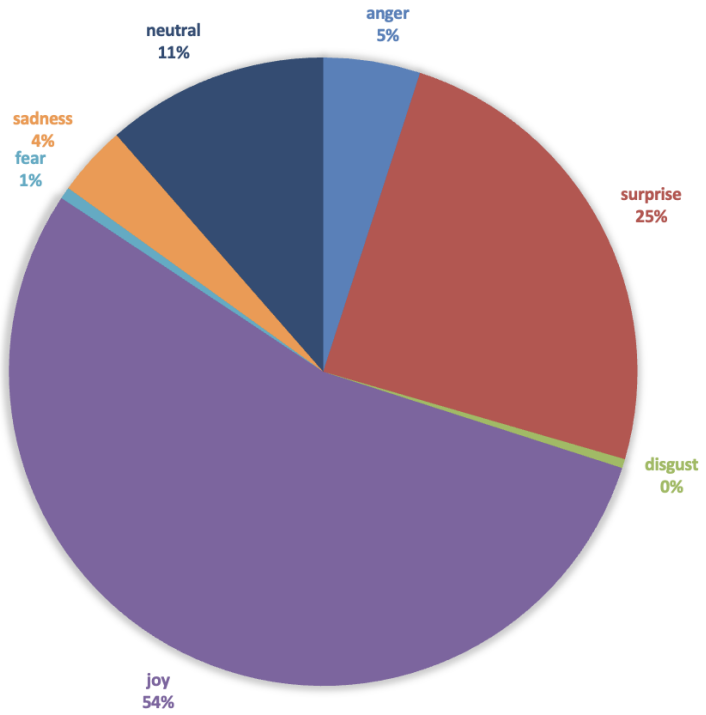

Figure 4.2: Pie chart showing the average expressed emotions in the comments.

Quite the same, the figure 4.2 shows that joy is also the dominant emotion expressed on average in the comments section, but has significantly less share of the total emotions pool with 54%, while many of the users express their surprise more than the content creators.

4.2 Visual Analysis

The raw data from the emotion detection phase was visualized as scatter plots to identify any clear indicators of correlations and to detect potential collinearity between different variables.

4.2.1 Emotions from the videos

As seen on the figure 4.3, the variable 'Anger' does not exhibit a strong correlation with Amazon ratings, with most instances showing low anger values below 0.2. An interesting observation is the increase in anger levels for Amazon ratings between 3.5 and 4.5, followed by a sharp drop as the ratings approach 5.

Figure 4.3: Scatter plot showing the variable 'anger' with respect to the amazon ratings.

Moving on to the variable 'Surprise,' we observe a slight increase in surprise levels as Amazon ratings approach 4 to 5, with some instances reaching values as high as 0.60. This indicates that products with higher Amazon ratings are more likely to evoke surprise in customers. Despite the small differences between anger and surprise, both emotions display dense clusters at low levels between Amazon ratings of 3 to 5. The plot can be seen on figure 4.4

The emotion 'Disgust' does not show a significant presence in most instances as seen on the figure 4.5, with outliers barely exceeding 0.20 in comparison to other variables in proportion. The limited vertical spread indicates that most values are clustered near 0, with very low variance.

Observing the variable 'Joy' reveals some interesting insights as shown

Figure 4.4: Scatter plot showing the variable 'surprise' with respect to the amazon ratings.

Figure 4.5: Scatter plot showing the variable 'disgust' with respect to the amazon ratings.

on figure 4.6, the most notable being that joy is a prominent emotion in the videos, consistently registering at high levels (above 0.60). Judging by the density, as Amazon ratings increase, the joy levels tend to remain more consistently high.

Looking at the variable 'Fear', we observe a pattern similar to that of

Figure 4.6: Scatter plot showing the variable 'joy' with respect to the amazon ratings.

the emotion 'Disgust', with a very low vertical spread and no significant or obvious linear relationship with Amazon ratings. The full plot can be seen on figure 4.7

Figure 4.7: Scatter plot showing the variable 'fear' with respect to the amazon ratings.

The variable 'Sadness' shows slight increases around an Amazon rating of 4, with a few outliers at this same value as seen on figure 4.8. However,

overall, sadness is one of the less prominent emotions. The low levels of sadness in lower ratings suggest that customers may express dissatisfaction through other emotions, such as anger, rather than sadness.

Figure 4.8: Scatter plot showing the variable 'Sadness' with respect to the amazon ratings.

The 'Neutral' state, representing an emotionless condition, shows almost no significant observations, with most instances remaining between levels 0 and 0.3. This is likely related to technical aspects discussed in video reviews, which often carry no emotional content, although the quantity of such reviews varies based on the product and the reviewer. The visualization can be seen on figure 4.9.

The only non-emotional variable, the number of Amazon reviews left for a product, is strongly correlated with Amazon ratings which can be clearly seen on figure 4.10. Products with higher ratings tend to attract more reviews. This is interesting because it suggests that many people prefer to express satisfaction for good products rather than dissatisfaction for poor ones.

Figure 4.9: Scatter plot showing the variable 'neutral' with respect to the amazon ratings.

Figure 4.10: Scatter plot showing the variable 'amazon reviews' with respect to the amazon ratings.

4.2.2 Emotions from the comments

As seen on the figure 4.11, the variable 'Anger' does exhibit a weak correlation with Amazon ratings, as the ratings get better, the levels of anger become more consistent at higher, with some outliers going to the levels of 0.6 to 0.8.

Figure 4.11: Scatter plot showing the variable 'anger' with respect to the amazon ratings.

The variable 'Surprise' leaves almost no obvious clues for finding linear relations, with most of the instances vertically spread over the same ratings and in quite large ranges, from 0.8 to as low as 0. The plot can be seen on figure 4.12. Surprise is one of the prominent emotions in the comments.

Figure 4.12: Scatter plot showing the variable 'surprise' with respect to the amazon ratings.

The emotion 'Disgust' again does not show a significant presence in most instances as seen on the figure 4.5, this is almost exactly the same as the emotions from the videos.

Figure 4.13: Scatter plot showing the variable 'disgust' with respect to the amazon ratings.

'Joy' is again one of the most prominent variables just like in the case of transcription emotions as shown on figure 4.6

Figure 4.14: Scatter plot showing the variable 'joy' with respect to the amazon ratings.

Looking at the variable 'Fear', we can see that around the rating of 4 there is a slight increase but other than that, there is no significant visual indicator of correlation. The full plot can be seen on figure 4.7

Figure 4.15: Scatter plot showing the variable 'fear' with respect to the amazon ratings.

The variable 'Sadness' shows slight and almost constant increases as the rating goes higher, as seen on figure 4.8. in contrast with the video emotions, sadness is more prominent in comments, showing that the commentors are more relaxed to express sadness than the content creators

The 'Neutral' state, exactly like the situation with the video emotions, is still not showing any significant correlations. The visualization can be seen on figure 4.9

The amazon review counts remains the same as previously mentioned and shown in figure 4.10.

4.3 Correlations and Collinearity

To determine if our variables can be used together effectively, we need to assess correlations and address collinearity, either by removing collinear variables or by combining them. Figure 4.18 presents the correlation ma-

Figure 4.16: Scatter plot showing the variable 'Sadness' with respect to the amazon ratings.

Figure 4.17: Scatter plot showing the variable 'neutral' with respect to the amazon ratings.

trix for the emotions in the videos, which reveals strong negative correlations between certain emotions, such as joy and anger, joy and surprise, and joy and sadness.

To further assess how these correlations might affect our analysis and to determine if they are significant enough to warrant additional fea-

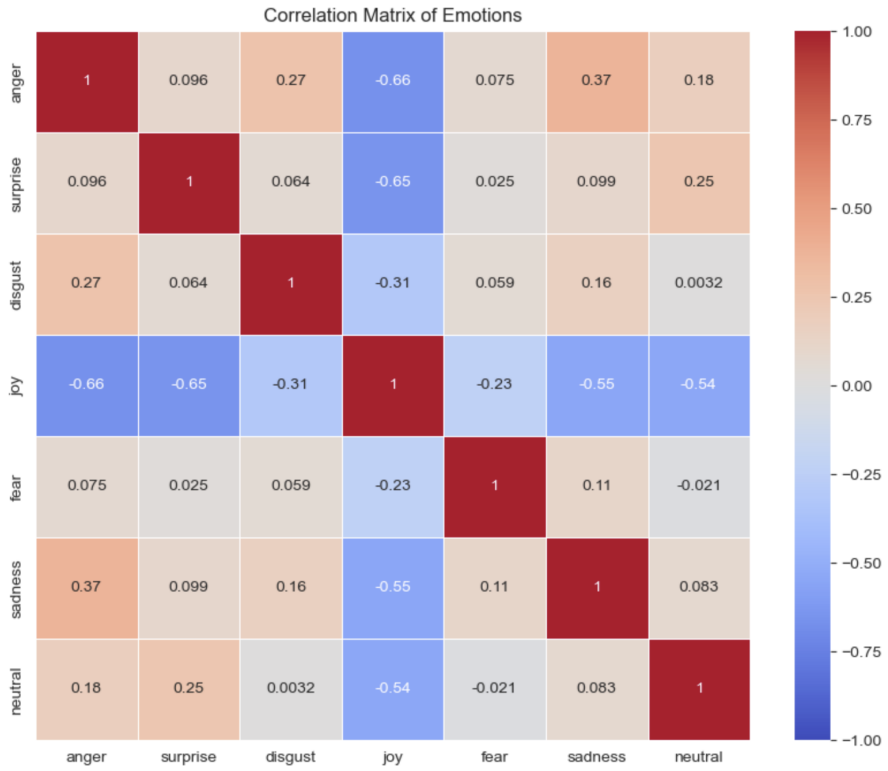

Figure 4.18: Correlation matrix, showing the correlation of all the videos' emotions with respect to each other.

ture engineering, we calculated the Variance Inflation Factor (VIF). The results are presented in Table 4.1.

Judging by the VIF scores, which are all below 4, there are no significant signs of collinearity among the variables, suggesting that they can be used as is. However, this does not guarantee their individual effectiveness.

We performed a similar analysis for the emotions in comments. The correlation matrix, shown in Figure 4.19, reveals a strong negative correlation between joy and surprise, and a strong positive correlation between neutral and surprise. Consequently, there is also a strong negative correlation between joy and neutral.

To further confirm the absence of multicollinearity, we calculated the VIF scores for the emotions in comments. As shown in Table 4.2, all VIF scores are below 4, indicating no significant multicollinearity issues, and

Feature	VIF
anger	2.557
surprise	2.097
disgust	1.325
joy	2.217
fear	1.102
sadness	2.032

Table 4.1: Variance Inflation Factor (VIF) for Emotions of videos

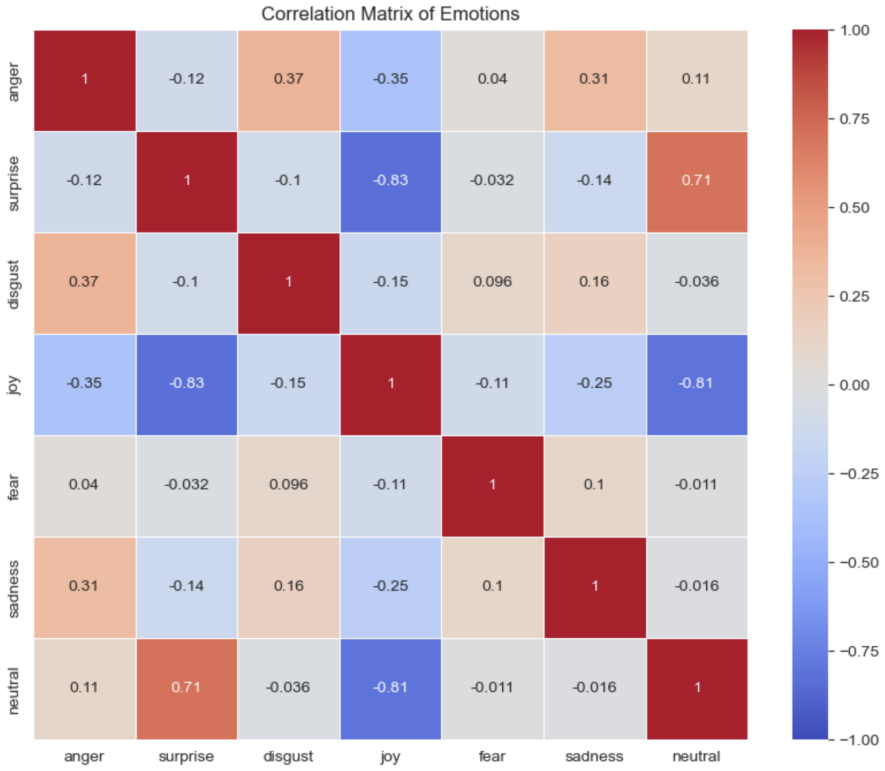

Figure 4.19: Correlation matrix, showing the correlation of all the comments' emotions with respect to each other.

thus the variables can be used without concern for multicollinearity.

4.4 Linear Model

Knowing that our variables do not exhibit signs of multicollinearity, we can proceed to fit a linear regression model to assess how well the model

Feature	VIF
anger	1.822
surprise	1.726
disgust	1.262
joy	1.739
fear	1.084
sadness	1.571

Table 4.2: Variance Inflation Factor (VIF) for Emotions of Comments

explains the variability in our data.

4.4.1 Video Emotions

The results for the model fitted using the emotions from the videos are presented in Table 4.4.1.

Variable	Coefficient	Std. Error	t-statistic	P-value
const	3.4952	0.135	25.919	0.000
anger	0.2137	0.214	0.998	0.318
surprise	0.0482	0.200	0.241	0.809
disgust	0.3231	0.541	0.597	0.550
joy	0.1234	0.152	0.815	0.415
fear	0.3156	0.328	0.963	0.336
sadness	-0.2069	0.234	-0.885	0.376
log(reviews)	0.0872	0.003	28.505	0.000
R-squared		0.222		
Adj. R-squared		0.220		
F-statistic		116.8		
Prob (F-stat)		3.62e-151		
AIC		1912		
BIC		1960		

Table 4.3: OLS Regression Results for Amazon Rating, fitted with videos' emotions

The intercept has a coefficient of 3.49, suggesting that the baseline Amazon rating for a product, given our samples, is approximately 3.49 when no other variables are considered.

The significant variable in our model is the number of Amazon reviews,

which has a coefficient of 0.08. This indicates that for every unit increase in the logarithm of Amazon review counts, the rating is expected to increase by 0.08 points.

The emotions in the video do not appear to be significant in our model, as none of the p-values for these variables are below the 0.05 threshold.

The R-squared value of our model indicates that approximately 22% of the variability in the data is explained by the included variables.

As seen in the partial regression plot for the intercept, the relationship between the residuals of the intercept and amazon ratings indicate a significant positive effect (figure4.20).

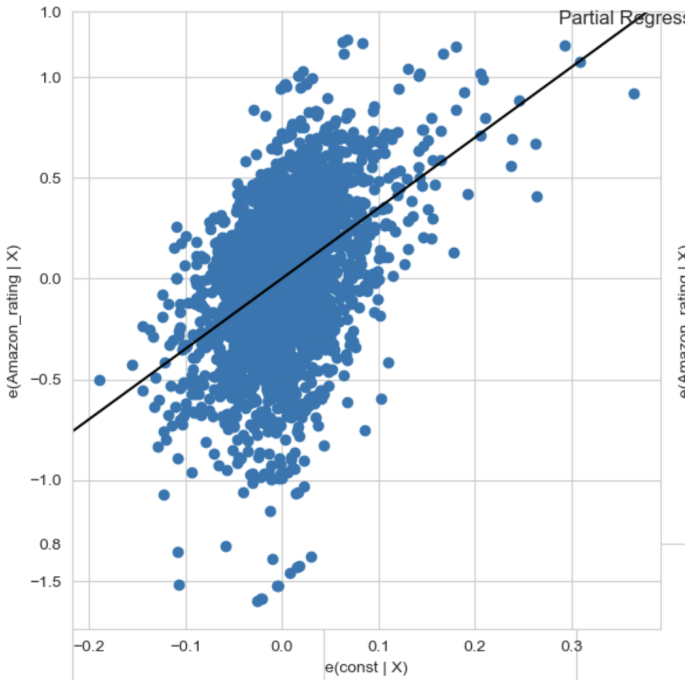

Figure 4.20: Partial Regression Plot for the Intercept (const) in Predicting Amazon Rating.

Again as seen in the figure4.21 and as expected from the results of the model, we observe a positive slope, indicating a positive relation between the amazon ratings and the amazon review counts.

Since the other variables are not of significance, all of them are almost or fully flat, not contributing to the amazon rating in a positive or nega-

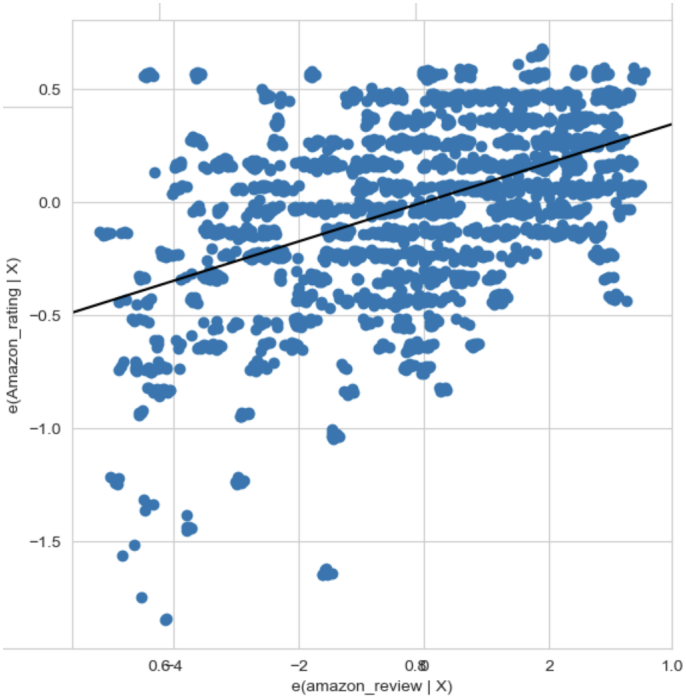

Figure 4.21: Partial Regression Plot for Amazon Reviews in Predicting Amazon Rating.

tive way. Anger which has the lowest p-value between the emotions has been shown as an example in figure4.22. Even though there is a slight positive slope observable, it is not significant enough to be taken into account, and this is the case for all the other variables.

4.4.2 Comment Emotions

Fitting a linear regression model using the emotions from the comments, resulted in a model with an R-squared of 27%, the full results are showed in the table4.4.2. This model improves over the previous model based on the r-squared.

The intercept has a coefficient of 3.77, suggesting that the baseline Amazon rating for a product is approximately 3.77 when no other variables are considered. The partial regression plot for the intercept can be seen in on figure4.23.

From the significant variables at level 0.05, **anger** has a negative co-

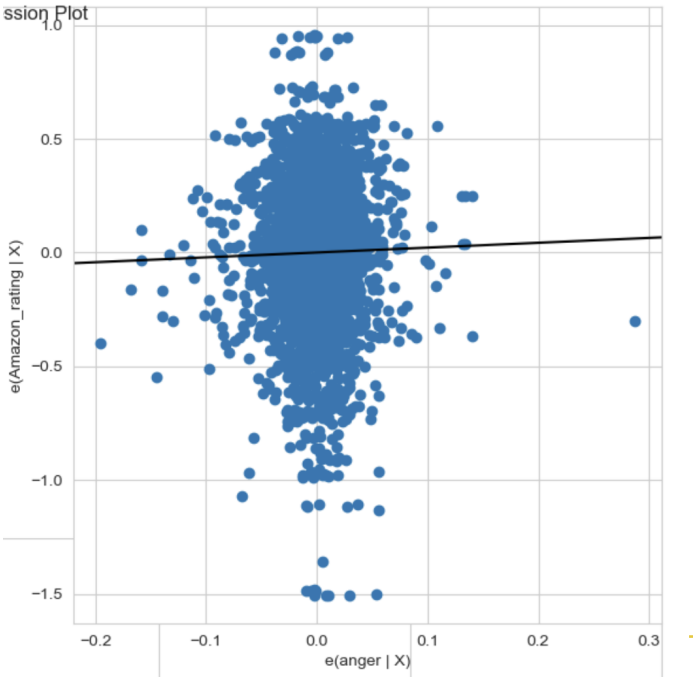

Figure 4.22: Partial Regression Plot for Anger in Predicting Amazon Rating.

efficient of -0.45, indicating that as anger increases, the Amazon rating tends to decrease, this is also demonstrated in the partial regression plot of the variable shown on figure 4.24.

Disgust also has a negative effect on the rating, with a coefficient of -1.06, meaning that more disgust leads to a lower rating. As seen on the figure4.25, the relationship is really weak since there is a quite wide spread observed.

Fear shows a strong negative relationship with a coefficient of -1.03, but again the same pattern which was observed with disgust can be seen on fear with varied vertical spread in its partial regression plot shown in figure 4.26 **Sadness** also contributes negatively to the rating, with a coefficient of -0.46. The partial regression plot for sadness as shown in figure 4.27, demonstrates a slight slope with more along-the-line spread instances than the previous two variables. The **Amazon review count** is the most significant positive predictor in the model, which has a co-

Table 4.4: OLS Regression Results for Amazon Rating

Variable	Coefficient	Std. Error	t-statistic	P-value
const	3.7671	0.160	23.523	0.000
anger	-0.4503	0.228	-1.976	0.048
surprise	-0.3098	0.217	-1.430	0.153
disgust	-1.0552	0.486	-2.169	0.030
joy	-0.1815	0.162	-1.117	0.264
fear	-1.0307	0.369	-2.790	0.005
sadness	-0.4618	0.221	-2.087	0.037
log(reviews)	0.0933	0.003	29.064	0.000
R-squared		0.270		
Adj. R-squared		0.268		
F-statistic		129.2		
Prob (F-stat)		5.51e-162		

Table 4.5: OLS Regression Results for Amazon Rating, fitted with comments' emotions

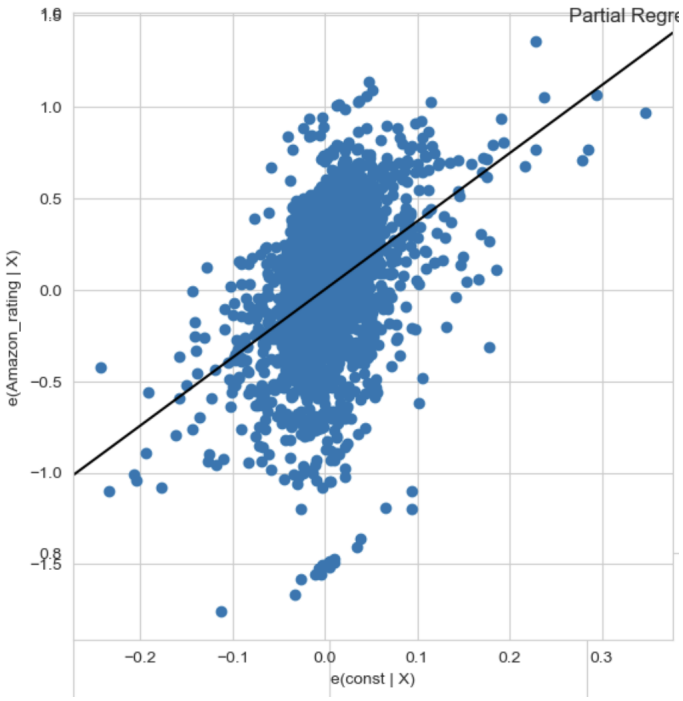

Figure 4.23: Partial Regression Plot for the Intercept in Predicting Amazon Rating.

efficient of 0.09. The partial regression plot shows an almost consistent positive correlation, demonstrated in figure 4.28

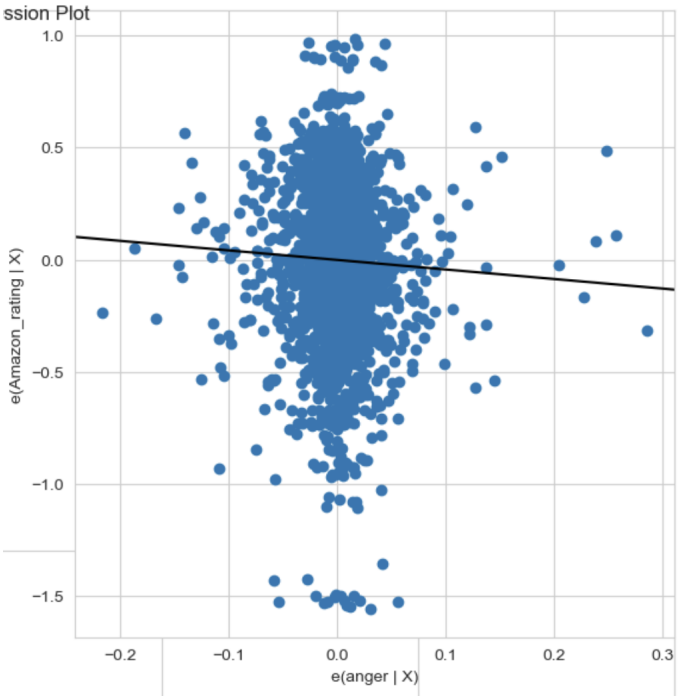

Figure 4.24: Partial Regression Plot for the Anger in Predicting Amazon Rating.

Surprise and Joy are the only non significant variables in our model with p-values over 0.05 level, also their partial regression plot is nearly flat, showing no significant effect on the output of amazon ratings.

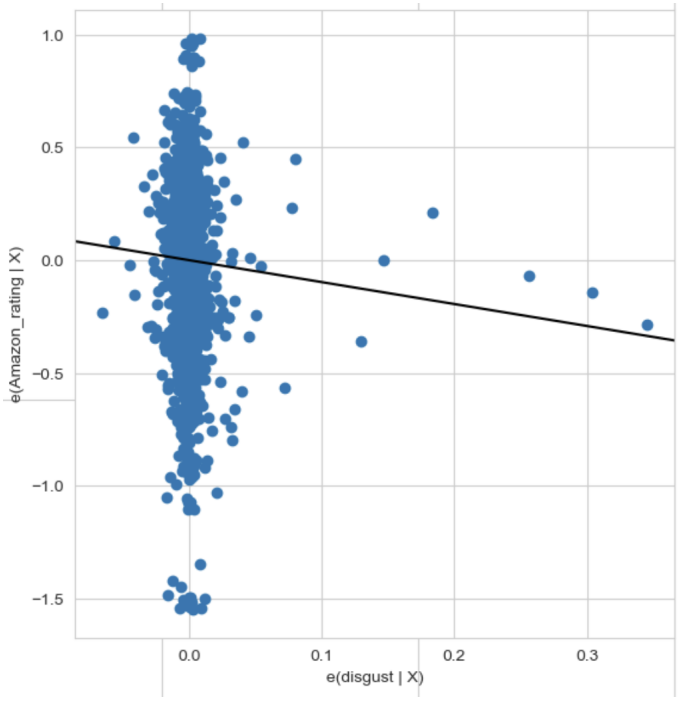

Figure 4.25: Partial Regression Plot for the disgust in Predicting Amazon Rating.

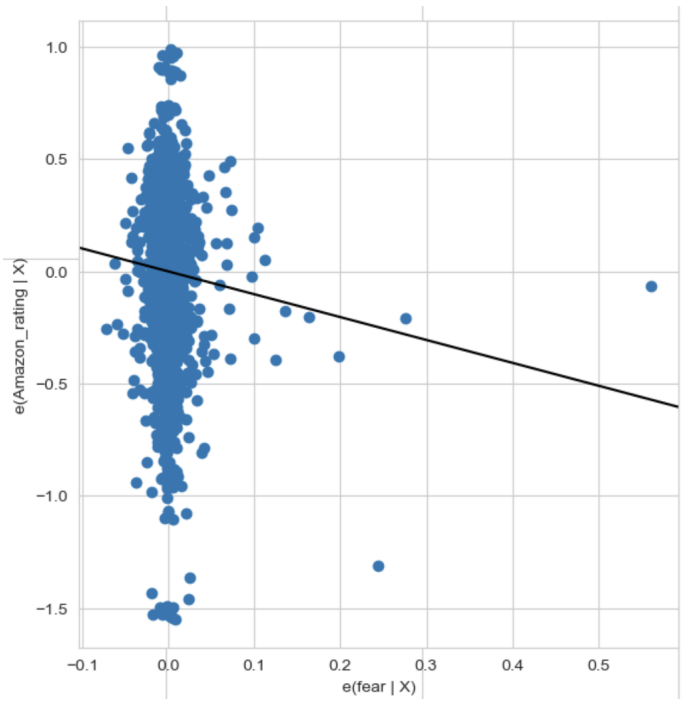

Figure 4.26: Partial Regression Plot for the Fear in Predicting Amazon Rating.

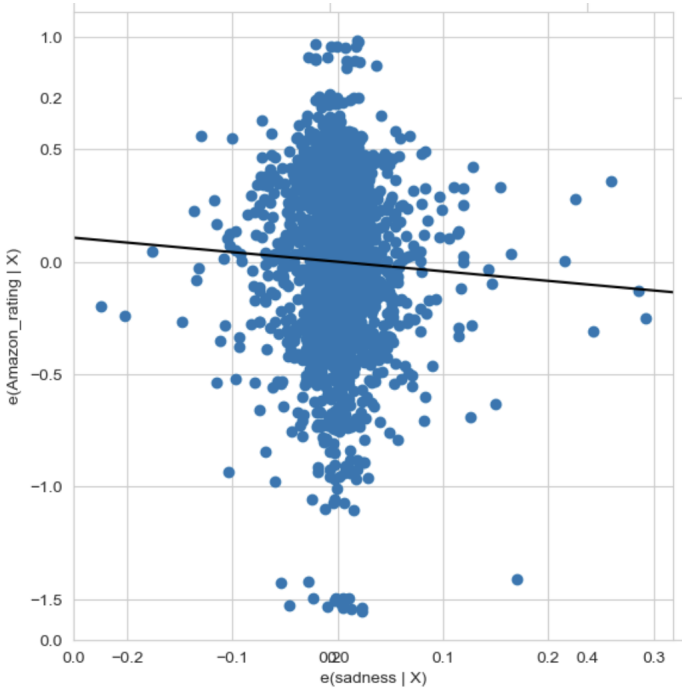

Figure 4.27: Partial Regression Plot for the Sadness in Predicting Amazon Rating.

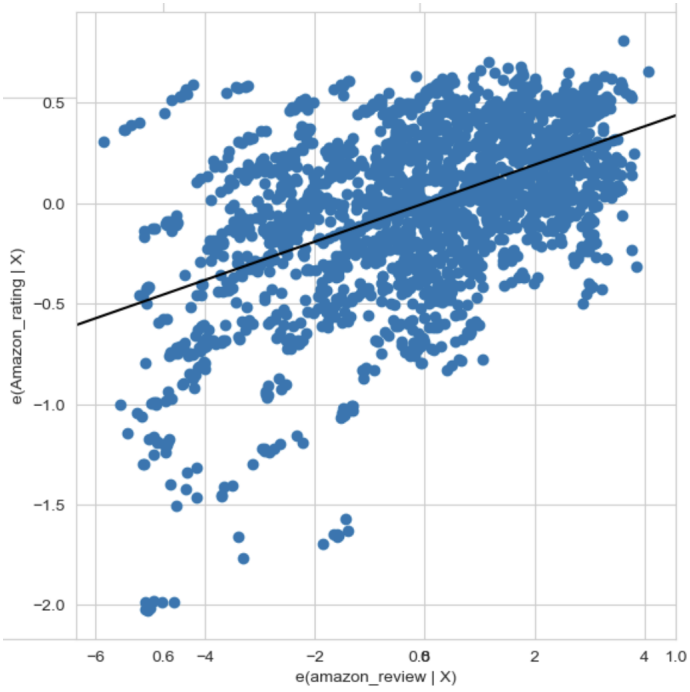

Figure 4.28: Partial Regression Plot for the Amazon review count in Predicting Amazon Rating.

Chapter 5

Discussion

It was shown during the study that a significant portion of our data consists emotionless and non expressive words, both in the video transcriptions and the comments, with them being constituted of 7.6% and 9.7% of adjectives respectively. The different percentage of emotion heavy words suggests that the content in the comments section is more expressive than what is recited in the videos by the content creators, potentially meaning that users are more likely to share their feeling-originated opinions of the products. The adjectives and the number of determiners used suggest that on average, less formal and more internet slang vocabulary and grammar are utilised in the comments section than in the content of the video.

In our analyses, the first key observation from the analyses is the role of certain emotions, particularly anger, happiness/joy, disgust, and fear, in predicting Amazon ratings. For both video reviews and comments, joy emerged as a consistent emotional presence, particularly in highly-rated products. This is consistent with prior research suggesting that positive emotional expressions, especially joy and admiration, are powerful drivers of favorable evaluations[58][62]. The strong presence of joy in high-rated products suggests that consumers tend to associate positive emotional experiences with high-quality products, which is likely a reflection of

their overall satisfaction with the product's performance.

Conversely, negative emotions such as anger, disgust, and fear were observed to have an inverse relationship with product ratings, with higher levels of these emotions were associated with lower product ratings. For example, anger was significantly correlated with lower ratings, particularly in comments, which is a key finding that aligns with previous studies on online reviews and sentiment analysis. Given that our work and Prahan[58] both agree that YouTube's sentiment is skewed toward positive emotions, we can see the connection of negative emotions and the ratings by exploring the work of Siersdorfer et al.[52] who examined similar trends in their study on YouTube comments, where they observed polarizing sentiments indicated polarizing content, and in Amazon's case, indicate products which their consumers have lower satisfaction and therefore they have potentially lower ratings. The negative influence of these emotions on ratings suggests that when consumers express dissatisfaction or frustration through anger or disgust, it often reflects as a poor experience with the product.

Interestingly, fear also demonstrated a negative correlation with ratings. Fear may be expressed in situations where the reviewer feels uncertain about a product's long-term reliability or safety. This pattern indicates that, while less frequent, fear can significantly influence perceptions of product quality, particularly for products where reliability or safety is a concern, such as electronics or high-end appliances.

The low VIFs for the emotions in both the comments and videos indicate that these emotions, whether they are related to the ratings, are independently contributing to the explanation of the target variable. This suggests that the emotions influencing the ratings do so without significant multicollinearity, meaning their individual impact on the ratings is not confounded by interdependencies with other emotional variables. Consequently, each emotion is able to explain variations in the ratings independently, ensuring that the relationship between emotions and ratings

is robust and not distorted by overlap between the explanatory variables.

The study also revealed notable differences in the emotional dynamics between YouTube video reviews and the corresponding comments. Emotions expressed in comments appeared to have a stronger correlation with Amazon ratings than those in the video transcriptions. This discrepancy might be explained by the more interactive nature of the comments section, where viewers not only react to the product but also engage with the opinions of the reviewer and other commenters. Commenters may be more inclined to express frustration or joy directly, resulting in a more emotionally charged environment compared to the more controlled emotional expression of the content creator in video reviews.

Additionally, this finding resonates with prior research that highlights the influence of consumer-to-consumer (C2C) interactions in shaping product evaluations. Penttinen et al.[43] emphasized the parasocial interaction between YouTubers and their audiences, where viewers develop one-sided relationships that may enhance their emotional engagement with the content. This could explain why emotions in comments have a stronger correlation with ratings; they reflect more immediate and reactive emotional responses compared to the more deliberate and potentially scripted nature of video reviews.

The OLS regression results provide insightful data, with the model for videos yielding an R-squared value of 22% and the model for comments achieving 27%. This indicates that the emotions expressed in the comments section account for 5% more variance in the ratings compared to those in the videos, suggesting that the emotions in comments better explain the ratings. The negligible difference between the adjusted R-squared and the R-squared values further suggests that the models are not overly complex, as the addition of predictors does not lead to substantial overfitting. Although the R-squared values are modest, they are evidence that emotions do contribute meaningfully to explaining the variability in ratings.

Our results show that emotions expressed in YouTube videos and comments of the videos are indeed correlated with amazon ratings, but these correlations differ in degree of influence and direction based on the origin of data, whether it was taken from the comments or the video, and the granularity of the emotions. They also show that commenters are more comfortable with sharing a diverse set of emotions that effectively discloses their opinion than the video producers. In the examinations we saw that the produced models, even though are statistically significant, still do not explain the whole variability of the data which means that emotions alone cannot be the deciding factor for explaining ratings, but can be regarded as very useful indicators which could be used among other variables to produce more accurate results.

Chapter 6

Conclusions

The aim of this study was to investigate the relationship between emotions expressed in YouTube product reviews and their respective comments and the corresponding Amazon product ratings. After analyzing more than 3,000 video transcriptions and the comments left on them, we discovered that a number of emotions, including as fear, disgust, rage, and delight, are significantly correlated with Amazon product evaluations. Joy was most commonly associated with higher ratings, reinforcing the idea that positive emotions often accompany consumer satisfaction. Conversely, negative emotions such as disgust and rage were associated with lower ratings, indicating that these emotions are usually used to convey dissatisfaction with a product.

One notable discovery is that product ratings were more strongly correlated with the emotions expressed in comments than in video reviews. This might be because comments are more direct and reactive, allowing viewers to express their thoughts and feelings more freely. The controlled environment of video reviews, on the other hand, may restrict content creators from displaying a full range of emotions.

While our results highlight the importance of emotional expression in predicting product ratings, the study also shows that emotions alone cannot fully explain ratings of the products. The modest R-squared

values from our regression models (22% for videos and 27% for comments) suggest that emotions are just one of the factors influencing consumer evaluations. Thus, while emotions can serve as useful indicators, they should be considered alongside other variables for a more comprehensive understanding of product ratings.

This research contributes to the growing field of sentiment and emotion analysis in consumer behavior, demonstrating the value of nuanced emotional categories over simple positive or negative sentiments. The findings also open avenues for further investigation into the role of emotions in shaping consumer perceptions, particularly in the context of online video reviews and comments.

6.1 Limitations and Challenges

Despite the significant insights gained from the study, there are several limitations that should be acknowledged. One primary limitation is the focus on text-based emotional analysis in a multimodal medium. While textual analysis provides a robust method for quantifying emotions, it overlooks other important non-verbal emotional cues, such as tone of voice, facial expressions, and body language. These cues, particularly in video reviews, may carry critical emotional information that cannot be captured through text alone. Incorporating multimodal analysis—which combines audio, visual, and textual data—would likely yield a more comprehensive understanding of the emotional dynamics in product reviews. This aligns with the recommendations of prior research, which suggests that sentiment analysis in online reviews could be enhanced by integrating prosodic and visual data alongside textual information.

Another limitation of this study lies in the scale and diversity of the dataset. Although the current dataset yields valuable insights, its scope is confined primarily to a not so large subset of product reviews. Expanding the dataset to encompass a broader array of products, user demographics,

and review contexts could improve the robustness and external validity of the results. Conducting experiments with larger and more diverse datasets would allow for more detailed analyses of how product types, reviewer characteristics, and situational factors influence emotional expressions in reviews. A more varied dataset would also facilitate the detection of nuanced emotional patterns that may not have emerged in the current study due to sample homogeneity.

Moreover, the current study focuses primarily on English-language reviews and comments. This language limitation could impact the generalizability of the findings, particularly given that emotions are expressed differently across cultures and languages. Future research could expand the scope to include multilingual datasets, allowing for a more globally representative analysis of online product reviews.

6.2 Future Research

The findings of this study have both theoretical and practical implications, on a theoretical level, this research contributes to the growing body of literature on emotion analysis in online reviews by providing empirical evidence of the emotional patterns associated with product ratings. It also highlights the importance of considering different emotional dynamics in reviews and comments, which can offer deeper insights into consumer sentiment.

Practically, these insights could be used in the development of automated review evaluation systems that leverage emotion detection to assess the reliability of online reviews. By identifying products that elicit disproportionate negative emotions relative to their ratings, such systems could potentially flag manipulated or deceptive ratings. This would be particularly valuable for e-commerce platforms like Amazon, where the prevalence of fake reviews continues to be a concern.

Another practical application of these findings lies in marketing and

product development. Understanding the emotional responses that consumers associate with different products can help businesses tailor their marketing strategies and product offerings to better meet consumer expectations. For instance, brands could focus on enhancing aspects of their products that evoke positive emotions like joy and admiration, while addressing concerns that trigger negative emotions such as fear or disgust.

Bibliography

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.

[3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692, 2019.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*, 2019.

[5] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. `https://huggingface.co/FacebookAI/roberta-base`, 2019. Accessed: 2024-09-23.

[6] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. GoEmotions: A dataset of fine-grained emotions. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online, July 2020. Association for Computational Linguistics.

[7] Sam Lowe. Roberta base goemotions model. https://huggingface.co/SamLowe/roberta-base-go_emotions, 2021. Accessed: 2024-09-23.

[8] Dezhi Yin, Samuel D. Bond, and Han Zhang. Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, 38(2):539–560, 2014.

[9] Armin Felbermayr and Alexandros Nanopoulos. The role of emotions for the perceived usefulness in online customer reviews. *Journal of Interactive Marketing*, 36(1):60–76, 2016.

[10] C. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, May 2014.

[11] Andrea Esuli and Fabrizio Sebastiani. SENTIWORDNET: A publicly available lexical resource for opinion mining. In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odijk, and Daniel Tapias, editors, *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May 2006. European Language Resources Association (ELRA).

[12] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61:2544–2558, 12 2010.

[13] Ditiman Hazarika, Gopal Konwar, Shuvam Deb, and Dibya Bora. Sentiment analysis on twitter by using textblob for natural language processing. pages 63–67, 01 2020.

[14] Jennifer S. Lerner and Dacher Keltner. Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, 14(4):473–493, 2000.

[15] Chethana Achar, Jane So, Nidhi Agrawal, and Adam Duhachek. What we feel and why we buy: the influence of emotions on consumer decision-making. *Current Opinion in Psychology*, 10:166–170, 2016. Consumer behavior.

[16] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon Press, New York, 1972.

[17] Google Developers. Youtube data api v3. https://developers.google.com/youtube/v3, 2024. Accessed: 2024-09-23.

[18] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv.*, 50(2), may 2017.

[19] Danish Raza and Reshma Nikhat. Purchasing pattern of consumers towards online and traditional shopping. 26:463–472, 01 2023.

[20] M.S.F. Sameeha and U.L. Milhana. A comparative study of traditional shopping and online shopping: Special reference to dharga town. *KALAM – International Journal Faculty of Arts and Culture, South Eastern University of Sri Lanka*, 14(4):1–2, 2021. Correspondence: milhanaul@seu.ac.lk.

[21] Christy M.K. Cheung and Dimple R. Thadani. The effectiveness of electronic word-of-mouth communication: A literature analysis. In *BLED 2010 Proceedings*. Association for Information Systems (AIS), 2010.

[22] Shaikh Abdul Hannan, Shaikh Jameel Ahmed, Quadri Naveed, and Rizwan Alam Thakur. Data mining and natural language processing methods for extracting opinions from customer reviews. *The International Journal of Intelligence Security and Public Affairs*, 3(6):52–58, 2012.

[23] Adam Hayes. Dotcom bubble. https://www.investopedia.com/terms/d/dotcom-bubble.asp, 2024. Accessed: 2024-09-23.

[24] R.L. Brandt. *One Click: Jeff Bezos and the Rise of Amazon.com*. Portfolio/Penguin, 2011.

[25] Amazon. Amazon marketplace: A winner for customers, sellers, and industry; new service grows over 200 percent in first four months, 2001. Accessed: 2024-08-21.

[26] Amazon. Fulfillment by amazon: How improving delivery fueled independent seller growth and success, 2024. Accessed: 2024-08-21.

[27] Amazon. Amazon global selling: Expand your business internationally. https://sell.amazon.com/global-selling. Accessed: 2024-08-21.

[28] Mengfan Zhai, Xinyue Wang, and Xijie Zhao. The importance of online customer reviews characteristics on remanufactured product sales: Evidence from

the mobile phone market on amazon.com. *Journal of Retailing and Consumer Services*, 77:103677, 2024.

[29] Georg Lackermair, Daniel Kailer, and Kenan Kanmaz. Importance of online product reviews from a consumer's perspective. *Advances in economics and business*, 1(1):1–5, 2013.

[30] Nasdaq. 92 billion reasons amazon can become the largest publicly traded company by 2027. *Nasdaq*, 2023. Accessed: 21-Aug-2024.

[31] Jawed Karim. Me at the zoo. YouTube Video, 2005. First video uploaded to YouTube.

[32] Ad Age Staff. Youtube: Fastest-growing website, 2006. Accessed: 2024-09-12.

[33] The Age. Google closes youtube deal, 2006. Accessed: 2024-09-12.

[34] Sydney Morning Herald Staff. Youtube serving up two billion videos daily, 2010. Accessed: 2024-09-12.

[35] Vincent Simonet. Classifying youtube channels: a practical system. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 1295–1304, New York, NY, USA, 2013. Association for Computing Machinery.

[36] YouTube Studio. Youtube studio. `https://studio.youtube.com/`. Accessed: 2024-09-12.

[37] YouTube. Press, 2023. Accessed: 2024-09-11.

[38] Dave Chaffey and Fiona Ellis-Chadwick. *Digital Marketing: Strategy, Implementation and Practice*. Pearson, Harlow, 5 edition, 2012.

[39] Eloise Aimee Aventajado. The influence of youtube on young consumers' purchase behavior. *International Journal of Modern Developments in Engineering and Science*, 2(1):31–38, Jan. 2023.

[40] Susie Khamis, Lawrence Ang, and Raymond Welling. Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, 8:1–18, 08 2016.

[41] Alexander Schouten, Loes Janssen, and Maegan Verspaget. *Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit*, pages 208–231. 03 2021.

[42] Rizwan Ahmed, Sumeet Seedani, Manoj Ahuja, and Sagar Paryani. Impact of celebrity endorsement on consumer buying behavior. *Journal of Marketing and Consumer Research*, 16:12–20, 11 2015.

[43] Valeria Penttinen, Robert Ciuchita, and Martina Čaić. Youtube it before you buy it: The role of parasocial interaction in consumer-to-consumer video reviews. *Journal of Interactive Marketing*, 57(4):561–582, 2022.

[44] Pantas H. Silaban, Wen-Kuo Chen, Tongam Sihol Nababan, Ixora Javanisa Eunike, and Andri Dayarana K. Silalahi. How travel vlogs on youtube influence consumer behavior: A use and gratification perspective and customer engagement. *Human Behavior and Emerging Technologies*, 2022:1–10, 2022. Citations: 21, First published: 20 June 2022, Academic Editor: Zheng Yan.

[45] A. Rohm and M. Weiss. *Herding Cats: A Strategic Approach to Social Media Marketing*. Digital and social media marketing and advertising collection. Business Expert Press, 2014.

[46] Google Support. User-generated content policy on youtube. `https://support.google.com/youtube/answer/3311596?hl=en`, 2023. Accessed: 2024-09-11.

[47] Cristos Goodrow. You know what's cool? a billion hours, 2017. Accessed: 2024-09-11.

[48] George A. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41, nov 1995.

[49] Soumya Sunkapaka and Nageshwar V. Sentiment analysis using telugu sentiwordnet. In *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, pages 1–5, 2023.

[50] Vijjini Anvesh Rao, Kaveri Anuranjana, and Radhika Mamidi. A sentiwordnet strategy for curriculum learning in sentiment analysis. In Elisabeth Métais, Farid Meziane, Helmut Horacek, and Philipp Cimiano, editors, *Natural Language Processing and Information Systems*, pages 170–178, Cham, 2020. Springer International Publishing.

[51] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference*

on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2017.

[52] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. How useful are your comments? analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 891–900, New York, NY, USA, 2010. Association for Computing Machinery.

[53] Pansy Nandwani and Rupali Verma. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):81, 2021.

[54] Olga Uryupina, Barbara Plank, Aliaksei Severyn, Agata Rotondi, and Alessandro Moschitti. SenTube: A corpus for sentiment analysis on YouTube social media. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4244–4249, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[55] Devang Shah and Masumee Parekh. From youtube comments to insights: A sentiment analysis of opinions on productivity tools. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 11(VIII):2302–2307, 2023.

[56] Sandipan Sahu, Raghvendra Kumar, Pathan MohdShafi, Jana Shafi, SeongKi Kim, and Muhammad Fazal Ijaz. A hybrid recommendation system of upcoming movies using sentiment analysis of youtube trailer reviews. *Mathematics*, 10(9), 2022.

[57] Feray Adigüzel. The effect of youtube reviews on video game sales. *Journal of Business Research - Turk*, 13:2096–2109, 09 2021.

[58] Rahul Pradhan. Extracting sentiments from youtube comments. In *2021 Sixth International Conference on Image Information Processing (ICIIP)*, volume 6, pages 1–4, 2021.

[59] Arghya Ray, Pradip Kumar Bala, and Rashmi Jain. Utilizing emotion scores for improving classifier performance for predicting customer's intended ratings from social media posts. *Benchmarking: An International Journal*, 28(2):438–464, 2021.

[60] Mohibullah Hawlader, Arjan Ghosh, Zaoyad Khan Raad, Wali Ahad Chowdhury, Md. Sazzad Hossain Shehan, and Faisal Bin Ashraf. Amazon product reviews: Sentiment analysis using supervised learning algorithms. In *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, pages 1–6, 2021.

[61] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[62] Sobia Wassan, Xi Chen, Tian Shen, Muhammad Waqar, and Noor Jhanjhi. Amazon product sentiment analysis using machine learning techniques. *REVISTA ARGENTINA DE CLINICA PSICOLOGICA*, 30:695–703, 03 2021.

[63] Mohammad Abu Kausar, Sallam Osman Fageeri, and Arockiasamy Soosaimanickam. Sentiment classification based on machine learning approaches in amazon product reviews. *Engineering, Technology amp; Applied Science Research*, 13(3):10849–10855, Jun. 2023.

[64] Chaehan So. What emotions make one or five stars? understanding ratings of online product reviews by sentiment analysis and xai. In Helmut Degen and Lauren Reinerman-Jones, editors, *Artificial Intelligence in HCI*, pages 412–421, Cham, 2020. Springer International Publishing.

[65] Mohammed Qorich and Rajae El Ouazzani. Text sentiment classification of amazon reviews using word embeddings and convolutional neural networks. *The Journal of Supercomputing*, 79(10):11029–11054, 2023.

[66] Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, Wentao Han, Minlie Huang, Qin Jin, Yanyan Lan, Yang Liu, Zhiyuan Liu, Zhiwu Lu, Xipeng Qiu, Ruihua Song, Jie Tang, Ji-Rong Wen, Jinhui Yuan, Wayne Xin Zhao, and Jun Zhu. Pre-trained models: Past, present and future. *AI Open*, 2:225–250, 2021.

[67] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *OpenAI*, 2018.

[68] István Üveges and Orsolya Ring. Hunembert: A fine-tuned bert-model for classifying sentiment and emotion in political communication. *IEEE Access*, 11:60267–60278, 2023.

[69] Xiangyu Qin, Zhiyu Wu, Tingting Zhang, Yanran Li, Jian Luan, Bin Wang, Li Wang, and Jinshi Cui. Bert-erc: fine-tuning bert is enough for emotion recognition in conversation. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023.

[70] Hashir Ali, Ehtesham Hashmi, Sule Yayilgan Yildirim, and Sarang Shaikh. Analyzing amazon products sentiment: A comparative study of machine and deep learning, and transformer-based techniques. *Electronics*, 13(7), 2024.

[71] Reddit User. LPT: When searching for items, paste "sort=review-count-rank" at the end of the hyperlink to rank items by number of reviews to truly see the more popular items., 2017. Accessed: 2024-09-13.

[72] Leonard Richardson. Beautiful soup 4. `https://www.crummy.com/software/BeautifulSoup/bs4/`, 2023. MIT License.

[73] ikaroskun. python-youtube: A python wrapper for youtube data api. `https://github.com/sns-sdks/python-youtube`, 2023. MIT License.

[74] The Pandas Development Team. pandas: Powerful data structures for data analysis, time series, and statistics. `https://pandas.pydata.org`, 2023. BSD 3-Clause License.

[75] Jonas Depoix. youtube-transcript-api: A python api for youtube video transcripts and subtitles. `https://github.com/jdepoix/youtube-transcript-api`, 2023.

[76] Google Cloud. Google cloud developer console. `https://console.cloud.google.com/cloud-resource-manager?pli=1`, 2024. Accessed: 2024.

[77] Xinyun Cheng, Xianglong Kong, Li Liao, and Bixin Li. A combined method for usage of nlp libraries towards analyzing software documents. In Schahram Dustdar, Eric Yu, Camille Salinesi, Dominique Rieu, and Vik Pant, editors, *Advanced Information Systems Engineering*, pages 515–529, Cham, 2020. Springer International Publishing.

Appendix

Python scripts

The following scripts were used to get Amazon data from the links and also to get youtube data from the video links.

```
def get_youtube_data(url, api_key = SecretKey):
    yt = Api(api_key=api_key)
    video = yt.get_video_by_id(video_id=getIDfromURL(url)).items[0]
    allComments = []
    video_title = video.snippet.title
    channel_title = video.snippet.channelTitle
    channel_id = video.snippet.channelId
    video_views = video.statistics.viewCount
    video_likes = video.statistics.likeCount
    total_comments = video.statistics.commentCount
    channel = yt.get_channel_info(channel_id=channel_id).items[0]
    subscriber_count = channel.statistics.subscriberCount
    channel_view = channel.statistics.viewCount
    transcription = YouTubeTranscriptApi.get_transcript(getIDfromURL(url),
      ↪ languages=['en'])
    try:
        comment_threads = yt.get_comment_threads(video_id=getIDfromURL(url),
          ↪ count=200, order='relevance').items
        for item in comment_threads:
            allComments.append({
                'comment': item.snippet.topLevelComment.snippet.textDisplay,
                'likes' : item.snippet.topLevelComment.snippet.likeCount,
                'replies' : item.snippet.totalReplyCount
            })
    except:
        comment_threads = []
    return {
        'YT_Video_Title' : video_title,
        'YT_Video_Views' : video_views,
        'YT_Video_Likes' : video_likes,
        'YT_Total_Comments' : total_comments,
        'YT_Comments' : allComments,
        'YT_Transcription' : transcription,
        'YT_Channel_Name' : channel_title,
        'YT_Subscriber_Count' : subscriber_count,
        'YT_Channel_View_Count' : channel_view
    }

def check_Video(id):
    try:
        transcript_list = YouTubeTranscriptApi.list_transcripts(id)
```

```
        yt = Api(api_key="AIzaSyDPHTM6ZlQM_-GKxWlEhAJIaAHw7bFMlbY")
        video = yt.get_video_by_id(video_id=id).items[0]
    except:
        return False
    x = str(transcript_list).find('English␣(auto-generated)')
    z = str(transcript_list).find('en␣("English")[TRANSLATABLE]')
    return x > 0 or z > 0
```

```
def get_youtube_data(url, api_key = SecretKey):
    yt = Api(api_key=api_key)
    video = yt.get_video_by_id(video_id=getIDfromURL(url)).items[0]
    allComments = []
    video_title = video.snippet.title
    channel_title = video.snippet.channelTitle
    channel_id = video.snippet.channelId
    video_views = video.statistics.viewCount
    video_likes = video.statistics.likeCount
    total_comments = video.statistics.commentCount
    channel = yt.get_channel_info(channel_id=channel_id).items[0]
    subscriber_count = channel.statistics.subscriberCount
    channel_view = channel.statistics.viewCount
    transcription = YouTubeTranscriptApi.get_transcript(getIDfromURL(url),
      ↪ languages=['en'])
    try:
        comment_threads = yt.get_comment_threads(video_id=getIDfromURL(url),
          ↪ count=200, order='relevance').items
        for item in comment_threads:
            allComments.append({
                'comment': item.snippet.topLevelComment.snippet.textDisplay,
                'likes' : item.snippet.topLevelComment.snippet.likeCount,
                'replies' : item.snippet.totalReplyCount
            })
    except:
        comment_threads = []
    return {
        'YT_Video_Title' : video_title,
        'YT_Video_Views' : video_views,
        'YT_Video_Likes' : video_likes,
        'YT_Total_Comments' : total_comments,
        'YT_Comments' : allComments,
        'YT_Transcription' : transcription,
        'YT_Channel_Name' : channel_title,
        'YT_Subscriber_Count' : subscriber_count,
        'YT_Channel_View_Count' : channel_view
    }

def check_Video(id):
    try:
        transcript_list = YouTubeTranscriptApi.list_transcripts(id)
        yt = Api(api_key="AIzaSyDPHTM6ZlQM_-GKxWlEhAJIaAHw7bFMlbY")
        video = yt.get_video_by_id(video_id=id).items[0]
    except:
        return False
    x = str(transcript_list).find('English␣(auto-generated)')
    z = str(transcript_list).find('en␣("English")[TRANSLATABLE]')
    return x > 0 or z > 0
```