NOVA IMS Information Management School

MDDM

Master's Degree Program in
Data-Driven Marketing

A Research Project on The Emotions Expressed by Users of Virtual Assistants Through Amazon Online Reviews

Vitória Fontana

Dissertation

presented as a partial requirement for obtaining the Master's Degree Program in Data-Driven Marketing

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

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

A RESEARCH PROJECT ON THE EMOTIONS EXPRESSED BY USERS OF VIRTUAL ASSISTANTS THROUGH AMAZON ONLINE REVIEWS

By/por

Vitória Fontana

Maser Thesis / Project Work presented as a partial requirement for obtaining the Master's degree in Data- Driven Marketing, with a specialization is Marketing Intelligence

Supervisor: Diego Costa Pinto

June 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used it. plagiarism or any form of undue use of information or falsification of results along the process, leading in its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Vitória Fontana

Lisbon, June 19th of 2023.

ABSTRACT

This project aims to examine whether Virtual Assistants have the ability to evoke emotions in users. The analysis specifically focuses on Amazon reviews on the product "Echo Dot Gen 4" extracted from their website. The project also incorporates relevant theories such as Social Agency and The Extended Self to provide supporting context. To effectively capture the sentiments of the users, a combined approach of Text Mining and Sentiment Analysis was employed. To determine the sentiment of individual words the AFFIN Lexicon and the NRC Lexicon are utilized. The findings of the analysis indicate that there is a significant positive effect of the average AFFIN sentiment score of a review on the average rating of the comments. With the NRC analysis it is possible to observe that most words have a positive sentiment and frequently contain the emotions joy and trust.

KEYWORDS

Virtual-Assistant-; Reviews; Sentiment-Analysis; Text-mining

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AFFIN The AFINN lexicon is a list of English terms manually rated for valence with an integer between -5 (negative) and +5 (positive). It was created by Finn Årup Nielsen between 2009 and 2011

AI Artificial Intelligence

NRC The NRC Emotion Intensity is a list of English words with real-valued scores of intensities for eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and additionally the “positive” and “negative” classification.

VA Virtual Assistant

1. INTRODUCTION

Artificial intelligence (AI) has seamlessly integrated into the daily lives of many individuals through various means such as mobile devices like cell phones, laptops, and tablets, as well as web-based services, applications, and websites. These devices are now well equipped with advanced artificial intelligence that possess human-like qualities and traits. This leads to the anthropomorphization of virtual assistants (VA) and chatbots, no longer viewed as mere tools, but rather as entities, transforming from assistants to companions (Costa, 2018). The first example of a digital VA available on a smartphone was Apple's "Siri," which was released as an application for iOS in February 2010 and later introduced as a feature on the iPhone 4S on October 4, 2011 (Afshar, 2021).

In 12 years, the AI tool gained popularity around the world. Based on the research carried out by Statista (2019), the authors forecast that the number of digital voice assistants could reach up to 8.4 billion units exceeding the current world population. In addition, the global smart speaker market could reach a value of US$15.6 billion by 2025 (Pathak et al., 2022).

VAs have gained worldwide recognition for their versatility in performing many tasks. These include setting reminders, providing information that would typically require a web search, controlling and monitoring the status of smart home devices such as lights, cameras, and thermostats, making and receiving phone calls, sending text messages among other functionalities. Apart from helping users for various purposes, virtual assistants possess social skills that enable them to offer social

companionship in online services. They have now become key in facilitating human-driven social interactions between service providers and customers, serving as the primary interfaces for customer-company interactions (Araujo, 2018).

After having an overview of the current situation of VAs and their popularity, some questions begin to emerge, such as which emotions users can demonstrate towards VAs and how they react after having a closer relationship with the device.

In recent years, several studies have been published on the topic. One notable example is from Kopp et al (2005) which describes an application, a conversational agent in a real-world setting. The system's design is described with a focus on how conversational behavior is accomplished. Recorded events from interactions between the conversational agent and museum visitors were analyzed for the kinds of dialogue people are willing to have with the agent.

In the study conducted by Reeves (1996), the author explores human interaction with computers and television programs, drawing parallels to human-to-human relationships. This psychological and sociological examination searches into the intricate dynamics of this relationship. The findings reveal that individuals tend to exhibit politeness towards computers and display differential treatment based on the gender of the computer's voice, with female-voiced computers eliciting distinct responses compared to their male counterparts.

There has been little analysis of Vas as modern servants and the gratitude expressed by users. This research aims to identify the emotions expressed by users of virtual assistants through online reviews. These reviews will be subjected to various

models of Sentiment Analysis, which allows to extract and quantify emotions and sentiments from textual data.

Important theories will be contemplated inside this research, such as Social Agency theory (2003) combined with the Extend Self theory (1988). According to the Social Agency Theory, when a computer exhibits characteristics like modulated intonation and human-like appearance, it enables individuals to perceive their interaction with the computer as a social encounter. Once this interaction is interpreted as a social interaction, the established social rules of communication between humans come into effect. Consequently, individuals tend to engage with the computer as if they were interacting with another human being. This phenomenon encourages deeper cognitive processing, as people strive to comprehend the artificial agent's communication and respond accordingly (Aeschlimann, 2020).

The Extended Self theory (1988) contributes in an equivalent way but assumes that individuals have distinct roles in different situations. In addition, people can change their behavior based on their current environment, for example how people act when specific factors and topics emerge. People need to adapt to fit in, at home, in the context of parenting, or in a team meeting, where everyone has different tasks and responsibilities. Individuals have multiple roles tailored to the unique situations they face (Mirbabaie et al., 2021).

The focus of this research lies in exploring the impact of Virtual Assistants on users' behavior. Given the fast accessibility of this technology, individuals may not have sufficient time to investigate the implications of having such technology in their homes.

This study shows how VAs can evoke emotions in users by performing an analysis through the utilization of new techniques like text mining and sentiment analysis.

The rest of the paper is organized as follows: chapter 2. offers an overview of the related literate on virtual assistants and discusses the development of the model and hypotheses; chapter 3 describes the dataset and data preprocessing steps as well as the used models; chapter 4 presents the results of the analysis; chapter 5. discusses the limitations and implications; chapter 6. concludes the main findings of the study.

2. LITERATURE REVIEW AND HYPOTHESES

Costa (2018) stated that AI tools often exhibit caregiving behaviors, even in interactions unrelated to direct assistance. These tools commonly employ maternal language, expressing concern and interest in the user's daily life and well-being. Consequently, they are recognized as empathetic and supportive entities that provide reassurance and care. The objective of this project is to investigate whether similar emotions can be identified in user comments through sentiment analysis. In a manner similar to Costa's (2018) paper, Wald et al. (2023) have also made a significant contribution by examining the usage of VAs within the family context. Unlike Costa (2018) whose focus is on maternal features of VAs, this study investigates how VAs can be accessed, operated, and enjoyed by multiple individuals concurrently, highlighting the role of modern technology.

Aeschlimann (2020) offers a distinct perspective on children's behavior towards voice assistants, posing the question if children can distinguish a voice between assistants and humans. The study employed both a human and a voice

assistant to provide guidance during a task. The results indicate that children hold distinct expectations for voice assistants when compared to humans. They do not anticipate voice assistants to mimic human behavior entirely. As a result, cooperation between humans and cooperation between humans and computers exhibit distinct characteristics. Moreover, the results indicate that children perceive artificial intelligence agents, such as Sila, as social beings capable of fulfilling roles like being a friend, providing comfort, and keeping secrets. Additionally, children attribute mental states to these agents, regarding them as intelligent entities with emotions.

Mirbabaie et al. (2021) examine the theories related to the virtual world, particularly emphasizing the concepts of "Social Identity" and the "Extended Self" theory. In this study VAs were incorporated as team members in an online work environment to explore how VAs influence the social identity and the extended self during virtual collaboration. However, while this research follows a similar approach as Mirbabaie et al. (2021), the main objective differs. Instead of analyzing the users' relationship with the VAs during actual communication, the analysis will be centered on the users' feedback, indirectly, like online reviews. Zhao (2020) examines a similar perspective, merging and synchronizing corporate and personal voice agents in one. A comparison of voice agents, one acting as a secretary and other as a housekeeper, were both designed and combined to exploit their different attributes. The results demonstrate that the merged VA performs based in terms of efficiency and satisfaction. Using one merged VA saves time and effort required by users, not needing to choose the VA for each service. The use of a merged VA leads to a more fluent information flow which results in higher efficiency and convenience for the user.

Additionally, users generally pay more attention to these two attributes, and thus it is not surprising that they prefer the simpler merged VA, compared to two separate ones for each task.

According to Kane et al. (2023), as more work transitions to online platforms and communication becomes predominantly text-based, the way emotions are conveyed online and their impact on others becomes increasingly significant. In online communication, verbalized emotional language replaces nonverbal emotional cues. Kane et al. (2023) revealed that the emotional language used by a partner during an interaction influenced the participant's positive and negative emotions afterwards. Moreover, participants who engaged with partners using more negative emotional language tended to respond with increased negativity themselves. The researchers identified an underlying mimicry mechanism that explained how the partner's language affected negative emotions, particularly when working with a partner of higher status. This finding is relevant to the current project's focus on analyzing Amazon online reviews, where individuals express their emotions about a VA tool. Understanding that people's emotions are influenced by various social interactions is crucial when examining the reviews, as these emotions may be affected by any other interactions.

Pang B et al. (2008) conduct an opinion mining and sentiment analysis which helped to establish sentiment analysis as an important research area. The paper discussed two main approaches to sentiment analysis: lexical-based methods and machine learning-based methods. The lexical-based approach uses the polarity of words to classify the emotions and the machine learning-based methods whereby

training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.

2.1 Sentiment Analysis, The Extended Self, Social Agency Theory

To analyze and quantify the emotions expressed in online reviews a sentiment analysis is utilized. Pang B et al. (2008) introduced the idea of using machine learning algorithms to classify text as positive, negative, or neutral based on the sentiment expressed in the text. However, in this project the lexical approach was used to determine the sentiment expressed on online reviews. The lexicon-based sentiment analysis is a widely used method for determining the emotional polarity of text, including positive, negative, or neutral sentiments. This technique relies on pre-established dictionaries containing words and phrases that are assigned to sentiment scores or categorized accordingly (Feldman, 2013).

According to Belk (1988), the extended self consists of three dimensions: the individual's body, their personal possessions, and their social relationships. The body is the most immediate and intimate aspect of the self, while possessions and social relationships serve as external expressions of one's identity. For example, the possessions we choose to own and the people we choose to associate with can provide insights into our preferences and beliefs. Also, the extended self can establish and maintain social identity and status.

Belk (2013) expanded on this theory in his book "Extended Self in a Digital World," exploring how digital technology has altered the concept of the extended self. In today's digital age, the author argues that virtual possessions, such as social media

profiles, online avatars, and digital files, have become increasingly important components of the extended self. Belk (2013) also shows the important implications for marketers and advertisers, who can use knowledge of an individual's extended self to tailor advertising messages and product offerings. Additionally, the extended self has significant consequences for how we interact with technology and the digital world. As we continue to incorporate digital technology into our lives, our concept of the extended self is likely to continue to evolve. In such cases, individuals tend to replicate traditional behaviors of a boss-servant relationship when interacting with VAs. For example, how customers behave with servers at a restaurant or cleaning professionals at home, but now with VAs, replacing human contact while continuing with the same roles. As a result, emotions that normally emerge towards another human now emerge in a VA.

Mayer (2003) presents social agency theory, another important theory regarding social behaviors. The theory builds upon the concepts of social identity theory and socialization theory to explain how individuals perceive and navigate their social environments. The theory proposes that individuals have both personal agency and social agency. Personal agency refers to an individual's ability to act independently and make decisions based on their own preferences and values, while social agency refers to an individual's ability to navigate and adapt to their social environment.

Social Agency Theory suggests that individuals use different strategies to balance personal and social agency, depending on the context and their goals. These strategies may include negotiation, compromise, and conformity to social norms and expectations, the theory also proposes that individuals are motivated to maintain a

positive social identity. The theory has been applied in various fields, including organizational behavior, social psychology, and political science. In organizational behavior, the theory suggests that employees may balance their desire for autonomy and control with the need to conform to organizational norms and expectations (Mayer, 2003). In political science, the theory has been used to understand how individuals balance their individual agency and social embeddedness when making political decisions. The theory suggests that organizations should create a supportive environment that allows individuals to balance their personal and social agency. It also highlights the importance of understanding how social identity and socialization influence individual behavior and decision-making.

Additionally, the Social Agency theory provides a framework for understanding human behavior in social situations and offers insights into how individuals navigate and adapt to their social environments. Social agency theory proposes that when computers exhibit social cues such as modulated intonation or a human-like appearance, people tend to interpret their interaction with the computer as being social in nature. This interpretation triggers the application of social rules that govern human-to-human communication, leading individuals to behave as if they were conversing with another human. Consequently, people are more inclined to engage in deep cognitive processing to understand the artificial agent's communication and respond accordingly (Aeschlimann 2020).

2.2 Conceptual Model and Hypotheses Development

The primary objective of this project is to examine the emotions that virtual assistants can evoke in humans by analyzing Amazon reviews. According to the Social Agency theory (2003), when computers exhibit social cues or possess human-like appearances, individuals perceive their interactions as social in nature. In addition, virtual assistants are designed with human-like features and tend to facilitate more relatable conversations.

In this study a sentiment analysis is employed to extract and analyze the emotional sentiments expressed in the Amazon reviews. Li and Wu (2010) used a similar approach by developing an algorithm to automatically assess the emotional polarity of textual data. This project followed a comparable methodology, utilizing R Studio to create a code and analyze emotions from words extracted from the Amazon reviews.

Öztürk and Ayvaz (2018) conducted a study comparing sentiments towards the Syrian refugee crisis using Turkish and English tweets. They employed a preprocessing step like the current project, the tweets were tokenized and converted all uppercase letters to lowercase, removing punctuation, links, and numbers from the datasets, and retaining only words for the analysis. After preprocessing, the tweets were attributed to sentiment scores and were calculated by examining the sentiment associated with each token. Finally, the sentiment scores of the tweets were aggregated based on the individual token sentiment scores.

In terms of the methodology employed, Fang (2015) adopted a similar approach in regards of tokenization. To facilitate the analysis, all the sentences were initially tokenized, meaning they were segmented into individual English words. This tokenization process allowed for a more granular examination of the text data. They focused on extracting subjective content, which encompassed all sentences expressing sentiment. A sentiment sentence was defined as a sentence that contained at least one positive or negative word.

Belk (2019) examines the relationships users develop with voice-controlled smart devices in his study titled "Servant, friend, or master? The relationships users build with voice-controlled smart devices." The study effectively portrays the user as the master and the device as the servant, capturing the dynamic between the two roles accurately.

The VA is designed to be subservient and lacks autonomy. Its main objective is to obey orders to the best of its ability but not always succeeding. When the device succeeds, both the user and the VA experience positive emotions such as happiness, satisfaction, and pride, fostering a lasting subordinate relationship. However, when the device fails, negative emotions arise, including anger, disappointment on the user's side, and a sense of frustration or sadness on the part of the VA, potentially leading to a breakdown or loss of the previously established trust in the relationship. If a dissatisfied user like this submits a review regarding the VA, it likely includes keywords of dissatisfaction and sadness.

One of the assumptions is that the relationship between humans and virtual assistants can resemble a boss and assistant relationship. In this case, the extracted words from the research should indicate some type of relationship or usage of the VA in work environment. Furthermore, a second assumption is that the relationship could also be perceived as a friendship, characterized by simpler and less formal language with casual conversation. This theory aligns with De Santis' (2019) findings, where the master and servant mode emerge when individuals take on the role of the master when interacting with a VA, as the device is already acting as “their servant”.

3. METHODS

This project aims to show consumers' emotions though their opinions about VAs found in online reviews. To achieve this, comments about Amazon’s smart speaker “echo dot” 4th gen were used, which are located in the comments section. To extract the reviews from Amazon’s website the “Amazon Reviews Exporter,” an extension of Google Chrome, was used and the data was exported to a csv file.

The dataset contains a total of 10,000 reviews and includes the following variables: "id" (i.e. the website's ID for user identification), "profile name” (i.e. the name of the user), "text" (i.e. the review's content), "date” (i.e. date including day, month and year) "title" (i.e. the user's title for the review), and "rating" (i.e. the number of stars given to a product, with a range of 1 to 5, where 1 indicates a negative review and 5 indicates a positive one). The data set was analyzed inside RStudio, which is an RStudio is an integrated development environment for R, a programming language for statistical computing and graphics.

In this study text mining is employed which describes the process of analyzing and extracting valuable inside from large, unstructured, and raw textual datasets, such as the reviews derived from Amazon's e-commerce. Using this data mining technique enables efficient collection and analysis of textual documents, commonly called corpus. Other examples of corpus that are frequently used for text mining are emails, social media posts or web pages.

To convert unstructured textual data into structured data and enable quantitative analysis and visualization natural language processing (NLP) techniques are implemented. Furthermore, a sentiment analysis is done to extract and quantify emotions and sentiments expressed by users in their reviews. Serrano-Guerrero (2015) highlights sentiment analysis, also known as opinion mining, as a prominent and contemporary research area in Information Processing. While textual information retrieval techniques primarily concentrate on processing, searching, or mining factual information, it is essential to acknowledge the existence of subjective components within text that express subjective characteristics alongside objective facts.

To enable sentiment analysis of unstructured data, several preprocessing steps are required. The raw corpus of Amazon reviews consists of strings of characters which also include punctuation, numbers, or other characters such as emojis. The first step is transforming this raw corpus into individual tokens as displayed in Figure I, also known as tokenization. The tokens can be individual units of text such as words but can also be n-grams which are combinations of multiple words that are commonly used together. According to Grefenstette (1994) to examine sentences, they need to be isolated, as most grammar focuses on sentence structures. To achieve sentence isolation, words

must be separated from the original stream of characters. Tokenization results in two types of tokens: one type represents units with identifiable character structures, such as punctuation, numbers, and dates, while the other type comprises units that will undergo morphological analysis.

Figure 1: The Process of Tokenization

After, all words beginning with capital letters were converted to lowercase. Lowercasing all words is a necessary step to normalize the text and ensure that words consisting of the same characters, but different capitalization are not treated as different words. Subsequently, numbers, blank spaces, and punctuation were removed because they are challenging to evaluate and do not contribute significant insights to sentiment analysis. The data should be formatted in a bag-of-words model, as proposed by Zhang (2010), and is widely recognized as a prominent technique for object categorization. Its fundamental concept involves mapping extracted key points to visual words and subsequently constructing an image representation through a histogram of these visual words.

The final pre-processing step and essential one is the removal of "stop words," which are frequently used words that serve as connectors in phrases but do not contribute meaningful value to the overall analysis. Silva (2003) defines stop words as words that possess a high frequency of occurrence in the data but hold low discrimination value. These words are often referred to as noise words. Stop words typically comprise the most used words in a language. When employing text mining techniques, it is a standard practice to remove these stop words from the text as they provide limited meaningful information. Common examples of stop words include 'a', 'the', 'to', 'for', 'but', among others.

Sentiment analysis involved the assignment of distinct sentiments to keywords, using a widely used sentiment lexicon called AFINN Sentiment Lexicon. This lexicon assigns scores to English words indicating positive and negative sentiment on a scale of –5 to +5. A sentiment score of lower than 0 indicates a word with a negative sentiment, while words with a score higher than 0 indicate a positive sentiment. Some reviews extracted use words with no sentiment attached, which are not covered in AFFINN lexicon library. Therefore, these reviews are not suited for this approach and are removed from the sample. The final sample size used in this study is 8,351.

In addition to providing sentiment scores using the AFFIN sentiment score, the project also utilized the NRC lexicon to categorize the words found within the reviews into different emotional buckets. This approach allowed for a more comprehensive analysis by assigning emotions to the words, thereby capturing the nuanced emotional content present in the reviews. Accordantly to Mohammad (2017) the NRC lexicon refers to a widely used sentiment lexicon developed by the National Research Council

of Canada. It is a comprehensive dictionary that contains a large collection of words and their associated sentiment scores across various emotional categories. The NRC lexicon provides information on different dimensions of sentiment, such as positive and negative sentiment, as well as eight specific emotions: joy, trust, anticipation, sadness, surprise, fear, anger, and disgust. Researchers and practitioners often rely on the NRC lexicon as a valuable resource for sentiment analysis tasks in natural language processing and computational linguistics.

4. RESULTS

The study utilized comments gathered from the reviews section of Amazon. The sample of reviews exported by "Amazon Reviews Exporter," an extension of Google Chrome, was randomly extracted. These comments underwent text mining and sentiment analysis processes to extract valuable insights. The 8,351 reviews analyzed were written between 2019 and 2023.

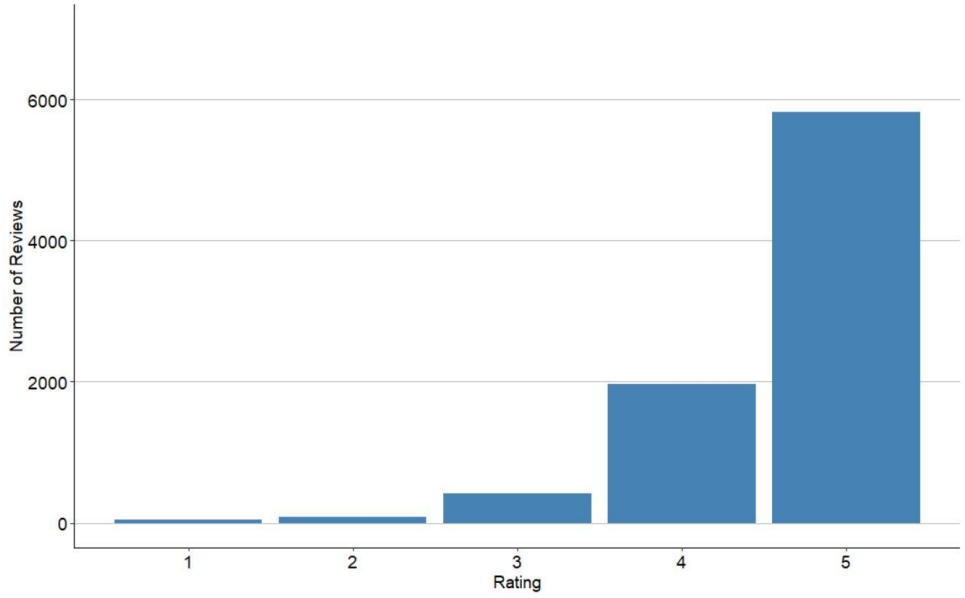
The distribution of reviews over the observation period is displayed on Table I. It can be observed that the number of reviews in 2019, is significantly lower compared to the other years. A possible explanation for that is the release date of Amazon's "echo dot" 4th gen which was released on the 25th of September in 2019. The average observed rating was similar over the years. The minimum average rating was in 2020 with 4.28 out of 5 and the maximum was in 2022 with 4.66, Which is also the year with the most reviews in the sample with a total of 6443 reviews. The biggest change can be observed from 2021 to 2022, when the rating increased from 4.29 to 4.66.

Table 1: Number of reviews per year

Year	Number of Reviews	Average Rating
2019	72	4.46
2020	286	4.28
2021	764	4.29
2022	6443	4.66
2023	786	4.62

The 5-start scale was the rating approach used by Amazon, a common scale system and one of the most used, begin used by big players, such as Google, Facebook and Yelp. The 5-star scale customers are asked to rate a product, service, or experience on a scale of 1 to 5 stars. The specific meanings of the 5-star rating scale labels may vary depending on the survey, but in general, 5 stars represent excellent or outstanding performance, while 1 point to a poor experience. For this specific case users could rate the product using this approach. When grouping the 8,351 reviews by the rating scale indicators it can be observed that most of the users rated the product with 5 stars. A total of 5,829 users rated Amazon's smart speaker with 5 stars representing 69.80% of the users in the sample. This information indicates that the product had incredibly good user satisfaction.

Figure 2: Number of Reviews Distributed by Rating Score



Upon removal of stop words, the analysis provided the most frequently occurring words within the sample. Table II provided valuable information, including the average rating associated with each word, the AFINN sentiment score, and the total frequency a word was used in the 8,351 reviews.

The findings revealed that the term "love" emerged as the most prominent word. It was mentioned 2,400 times in reviews, with an average rating of 4.80, when the term was used. This indicates a genuine expression of admiration from users towards the product. Additionally, positive words such as "Like," "Good," and "Great" appeared in the top 3, further supporting the initial argument. Notably, the top 15 most frequently used words all have a positive AFFIN sentiment score, being higher than 0.

Table 2: Most Frequent Words by AFFIN Sentiment Score

	Word	Average Rating	Sentiment Score	Frequency
1	love	4.80	3	2400
2	like	4.33	2	2336
3	great	4.68	3	2325
4	good	4.40	3	1689
5	easy	4.71	1	1158
6	better	4.25	2	1128
7	smart	4.45	1	1011
8	want	4.14	1	775
9	nice	4.51	3	558
10	pretty	4.34	1	447
11	recommend	4.65	2	411
12	fun	4.56	4	372
13	help	4.33	2	364
14	happy	4.68	3	359
15	amazing	4.82	4	352

Figure III illustrates a word cloud visualization depicting the most frequently used words in a review. In this visualization, the size and position of each word correspond to its frequency or count within the review. It is possible to see words like “Love”, “Like” and “Great” taking a more central place in the overall distribution and being bigger in size.

Figure 3: Word Cloud of Most Frequent Words by AFFIN Sentiment Score

In comparison, Table III shows the top 15 frequently occurring negative words, which are the ones with AFINN score values lower than 0. Some words such as "Alarm", which can have diverse applications in different contexts, had an average rating of 4.43. This rating is unexpectedly high for a word with a negative AFFIN sentiment score, where a lower rating would be more consistent. Similarly, the word "Alone" had an average rating of 4.55, defying its negative connotation. Even the word "Difficult" which carries an intense sense of negativity, had an average rating of 4.04 when used in a review.

Table 3: Most Frequent Words with a Negative AFFIN Sentiment Score

	Word	Average Rating	Sentiment Score	Frequency
1	alarm	4.43	-2	372
2	problem	3.86	-2	310
3	hard	4.14	-1	194
4	bad	4.01	-3	165
5	fire	4.26	-2	157
6	stop	3.60	-1	156
7	wrong	4.07	-2	147
8	drop	4.36	-1	146
9	problems	3.91	-2	144
10	disappointed	3.80	-2	134
11	annoying	3.68	-2	128
12	trouble	3.99	-2	101
13	pay	3.86	-1	90
14	alone	4.55	-2	83
15	difficult	4.04	-1	79

Figure IV illustrates a word cloud visualization depicting the most frequently used words in a bad review. It is possible to see words like “Alarm”, “Problem” and “Hard” taking a more central place in the overall distribution and being bigger in size.

Figure 4: Word cloud of most frequent words with a negative AFFIN sentiment score

As seen in Table III, the top 15 most used words with negative AFFIN sentiment scores had been used in reviews with higher ratings. To examine whether the words and their respective AFFIN sentiment score are consistent with the rating of the review, the 5-star reviews and the 1-2 start reviews are considered for the following analysis.

When considering only words that are used in 5-star reviews in Table IV, it is possible to see similar words as in Table II. This can be explained by the share of 5-star reviews in the sample since they account for 69,80% of the total amount of reviews. Table IV shows that the AFFIN sentiment score of the most frequently used words in 5-star reviews is consistent with the rating. The top 13 words all have an AFFIN sentiment score of 1 or higher, indicating words with a more positive sentiment.

Table 4: Most Frequent Words Used in 5-star Reviews

	Word	Average Rating	Sentiment Score	Frequency
1	love	5	3	2007
2	great	5	3	1738
3	like	5	2	1303
4	good	5	3	975
5	easy	5	1	883
6	smart	5	1	669
7	better	5	2	565
8	want	5	1	404
9	nice	5	3	360
10	recommend	5	2	312
11	amazing	5	4	305
12	fun	5	4	283
13	happy	5	3	270
14	alarm	5	-2	261
15	best	5	3	252

When analyzing the top 15 most frequent words in 1 and 2-star reviews, displayed in Table V, a difference from Table III can be observed, which shows the top 15 most frequent negative words. Unlike in the previous table, the rating and the sentiment of the words used are not consistent. The majority of the most frequently used words in 1- and 2-star reviews have a positive AFFIN sentiment score and only four words, “Problem”, “Stop”, “Alarm” and “Worse” have negative sentiment scores. In addition, words like “Good”, “Like” and “Fun” which have an AFFIN sentiment score of +2, are frequently used in 1 and 2-star reviews, which shows that these words with positive sentiment can be used in bad-rated reviews.

A possible explanation for this output might be that people use more kind words in bad reviews to soften their tone. The same can be done by Amazon's website, to remove bad reviews that use inappropriate language, leaving only the politest ones. For example, words with an AFFIN sentiment score of –5, that can be described as offensive, can only be found once in the sample. To support the first assumption, where users used good words in reviews with bad ratings, Table V shows words with a positive AFFIN score begin used in 1- and 2-star reviews.

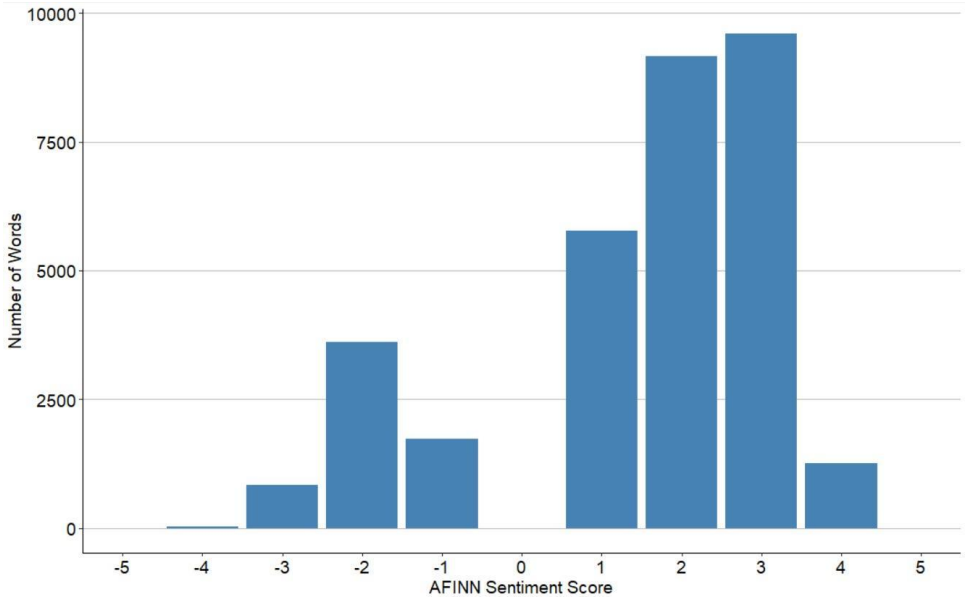
Table 5: Most Frequent Words Used in 1- and 2- star Reviews

	Word	Average Rating	Sentiment Score	Frequency
1	like	1.42	2	133
2	want	1.45	1	85
3	good	1.61	3	74
4	better	1.58	2	65
5	smart	1.43	1	61
6	problem	1.43	-2	40
7	support	1.32	2	38
8	stop	1.38	-1	37
9	alarm	1.71	-2	31
10	great	1.68	3	31
11	help	1.42	2	31
12	worse	1.48	-3	27
13	fine	1.54	2	26
14	pretty	1.77	1	26
15	fun	1.22	4	23

To gain a better understanding of the reviews it is helpful to analyze the distribution of words based on their sentiment score. According to the AFINN lexicon

there is only one expression assigned to the sentiment score equal to 0 which was not used in the sample. It is possible to see that the biggest number of words are defined as positive ones with a sentiment score of +2 and +3, which shows that the product has a good adoption, and the users express good feelings towards the same.

Figure 5: Distribution of Words Based on the AFFINN Sentiment Score

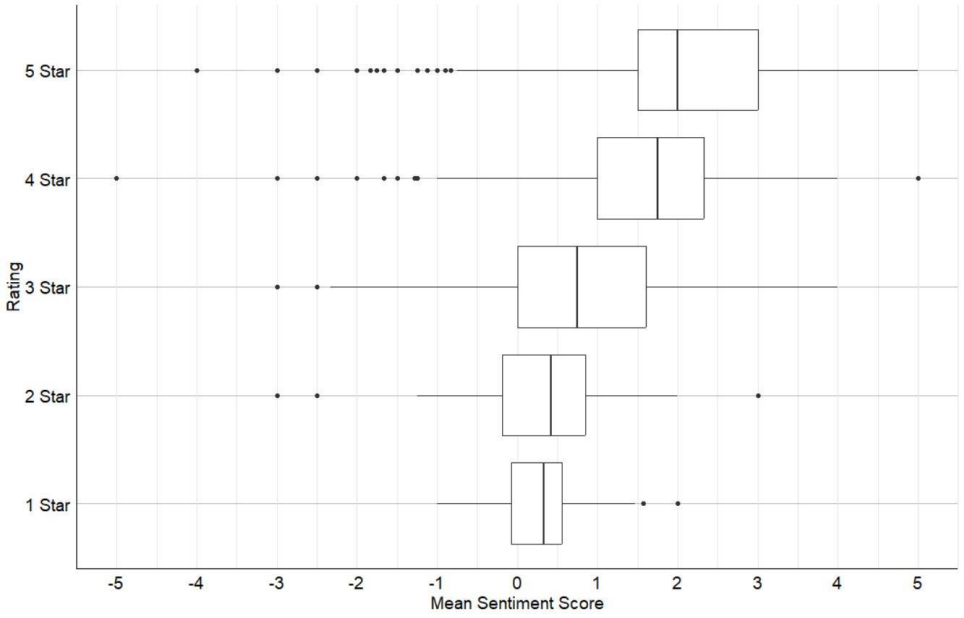


When examining the relationship between two variables it is a common approach to analyze their linear correlation. The term correlation commonly refers to the association between two variables, aiming to quantify this relationship and determine its strength. In this case, a weak positive linear correlation of 0.3297 between the ratings and the average AFFIN sentiment score of all the words used in the reviews can be observed.

To further investigate the relationship of the two variables, Figure VI provides a boxplot illustrating the relationship between the two variables graphically. First of all, it is notable that the median value of the average AFFIN sentiment score increases for

higher ratings. This might suggest that there is a positive effect of higher sentiment scores on the rating. The greater the use of positive words, the higher the ratings tend to be. However, as previously shown positive words are not exclusively used in good reviews but can also be found in negative reviews. This becomes evident when considering Table III which shows the most frequently used words in bad reviews. In addition, in figure VI various outliers can be observed, where users give 4- and 5-star but use predominantly words with a negative sentiment.

Figure 6: Correlation Between Ratings and AFFINN Sentiment Score



To quantify the relationship between the average AFFIN sentiment score of each review and the corresponding rating linear regression was conducted. As displayed on Table VI, a review with a neutral average sentiment score has an average rating of 4.260 indicated by the estimate of the constant. An increase of the average AFINN sentiment score by one led on average to an improvement of the rating by

0.192. The effect of the average AFINN sentiment score on the rating is positive and significant at the 1% significance level.

Furthermore, it is important to consider the coefficient of determination (R^2) which is a statistical measure that indicates the share of the variation of the dependent variable that can be explained by the model. For this case, 10.9% of the variation of the rating can be explained by mean sentiment score of a review.

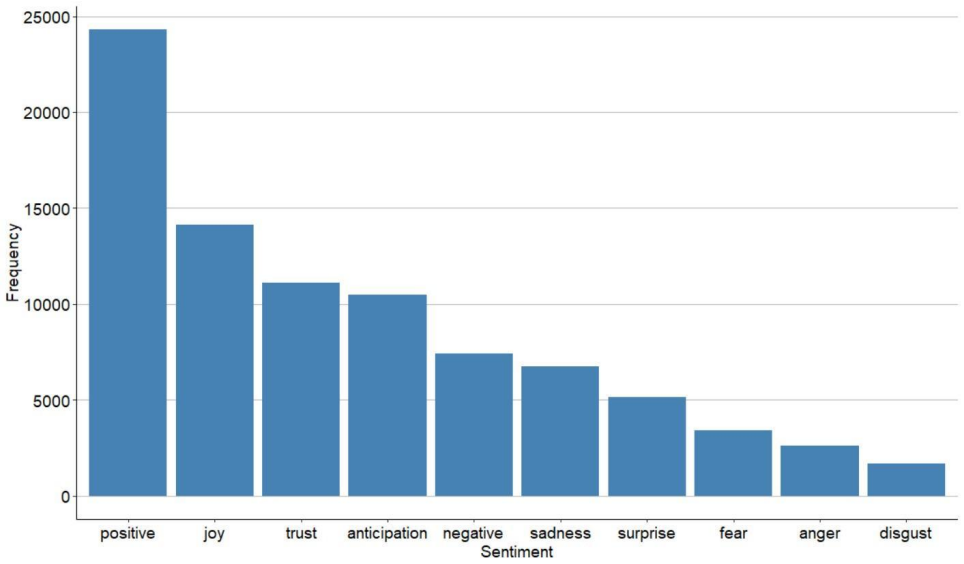
Table 6: Regression Results

	Rating
Mean Sentiment Score	0.192***
	(0.006)
Constant	4.260***
	(0.013)
Observations	8,351
R^2	0.109
Adjusted R^2	0.109
Residual Std. Error	0.651 (df = 8349)
F Statistic	1,018.283*** (df = 1; 8349)
Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

The first hypothesis can be confirmed, VAs can evoke sentiments in humans. As the results were analyzed a relationship between the ratings and the average AFFIN score could be seen. The AFFIN sentiment score can resemble the users' reviews as a whole and the rating as the output of this interaction.

Instead of assigning a numerical value to a word, as in the AFFINN lexicon, the NRC lexicon assigns them to two sentiments, positive and negative, but also to pre-defined eight emotions. It is possible to see that “positive” is the predominant sentiment with a higher count of words than any other sentiment or emotion. The most frequent emotion is “joy”, followed by “trust”.

Figure 7: Words distributed by NRC lexicon



One of the underlying assumptions is that the interaction between humans and VAs can resemble a hierarchical relationship, akin to a boss and assistant dynamic. In this context, the words extracted from the research are expected to reflect expressions of authority originating from the users. By employing the NRC lexicon, the study observed that the category of "trust" ranked second most frequent emotion. Although the concept of a work environment may not be directly associated with this emotion, the techniques utilized enabled the research to approach the closest explanation for a potential boss-assistant relationship.

Additionally, another assumption is that the relationship between humans and virtual assistants could also be perceived as a friendship, characterized by informal and casual language akin to casual conversations. By employing the NRC lexicon once again, it became evident that "joy" emerged as the most prominent emotion within the sample. While a relationship of friendship cannot be entirely linked to the "joy" alone, it does indicate that this particular sentiment was evoked in users, potentially resembling a friendly relationship with the VA.

5. DISCUSSION

When analyzing the dataset consisting of Amazon's user reviews for the smart speaker, model "echo dot" 4th gen, it stands out that the ratings are not distributed equally among the five rating options. Out of 8,351 reviews in the sample 69,80% are 5-star ratings. When considering both 5-star and 4-star ratings, they account for a total of 93,33% of the ratings. An explanation for this distribution of the ratings could be due to the elevated level of customer satisfaction with the device. Accordingly to Brill et al. (2019) in the research called "Siri, Alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications" results confirmed that expectations and confirmation of expectations have a positive and significant relationship on customer satisfaction with digital assistants. From a practice perspective, the study provided evidence that customer expectations are being satisfied through the digital assistant interaction experience.

However, another possible explanation for the high number of positive reviews within the sample is the fact that comments can be considered as a vitrine for

products. Firstly, due to the expansive reach of the internet and the virtually unlimited online space available to e-tailers, the number and variety of products offered online have grown exponentially. Consequently, online consumers often face the challenge of lacking sufficient knowledge and time to make optimal decisions among numerous competing products found across various websites.

Online user review systems serve as platforms for consumers to share their opinions and experiences regarding products. Inside Amazon's website the reviews section users go specifically to leave comments related to one specific product, the website owns the product, so these reviews are under a controlled environment, where the website can keep the highest rated ones and remove others. As stated by Duan et al. (2008), reviews are considered a digitalized word of mouth and have a significant impact on product sales and consumer decision-making. The abundance of information on the internet has resulted in information overload among online users, as highlighted by Cao et al. (2011).

In the context of online reviews, the sequential bias refers to the influence of previous, existing reviews on subsequent reviews. Chauhan et al. (2011) demonstrate the existence of such sequential bias in online reviews for search and experience goods and their dependency from previous review and reviewer characteristics.

Another method to validate if the reviews, could be based on another platform where users leave comments and compare results and satisfaction rates, like twitter, where unfiltered comments can be found with no specific subject, only the users' options in an open environment, this approach was not applied in this project

due to the restrictions recently implemented on the platform, that make the extraction of tweets using Twitter’s Api not possible.

Another aspect to consider is the absence of other VAs, such as Google's assistant, being sold on Amazon's website, which could serve as a point of comparison. Another important point to note is that when conducting text mining, personal information such as gender, age, occupation, and country is not taken into account. Therefore, there is no input available regarding the users' personal information that could provide further insights or analysis in relation to these factors.

6. CONCLUSION

The objective of the project is to extract emotions from Amazon reviews and demonstrate how they influence user behavior. During the analysis of the most frequent words in positive and negative reviews, users consistently used positive words in positive reviews. However, no consistent choice of words could be observed in negative reviews as many words with a positive sentiment score were used. This suggests that dissatisfied users may choose to express their emotions more politely. Through the analysis, it was observed that an increase of the average AFINN sentiment score of a review has a positive and significant effect on the rating. This indicates that users' sentiments and emotions played a crucial role in shaping their evaluations and perceptions of products.

By implementing the NRC lexicon one of the underlying assumptions about humans and VAs resembling a hierarchical relationship hints at a possible explanation. The study observed that the category of "trust" was ranked second most frequent

emotion, the concept of a work environment may not be directly associated with this emotion, so the servant-boss relationship would resemble more to a trust one.

The same was applied to another assumption, that the relationship between humans and virtual assistants could also be perceived as a friendship. The NRC lexicon categorized the emotion of "joy" as the most prominent emotion within the sample. A relationship of friendship cannot be entirely linked to the "joy" alone, but it does indicate that this sentiment was seen in users, potentially resembling a friendly relationship with the VA.

REFERENCES

Afshar, V. (2021, April 7). AI-powered virtual assistants and the future of work. ZDNET.https://www.zdnet.com/article/ai-powered-virtual-assistants-and-future-of-work/

Aeschlimann, S., Bleiker, M., Wechner, M., & Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. Computers in Human Behavior, 112, 106466.

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Computers in Human Behavior, 85, 183-189.

Belk, R. W. (1988). Possessions and the extended self. Journal of consumer research, 15(2), 139-168.

Belk, Russell W. (1988). Possessions and the Extended Self. Journal of Consumer Research, Volume 15, 139–168.

Boudon, R. (2016). The unintended consequences of social action. Springer.

Brill, T. M., Munoz, L., & Miller, R. J. (2019). Siri, Alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications. Journal of Marketing Management, 35(15-16), 1401-1436.

Costa, P. (2018). Conversing with personal digital assistants: On gender and artificial intelligence. Journal of Science and Technology of the Arts, 10(3), 59-72.

De Santis, E. (2019). Smart objects and human beings: a study on master servant relationships' patterns.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 1-14.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.

Grefenstette. G. and Tapanainen, P. (1994) What Is a Word, What Is a Sentence? Problems of Tokenization. The 3rd International Conference on Computational Lexicography, Budapest, 7-10 July 1994, 79-87.

Haque, T. U., Saber, N. N., & Shah, F. M. (2018, May). Sentiment analysis on large scale Amazon product reviews. In 2018 IEEE international conference on innovative research and development (ICIRD) (pp. 1-6). IEEE.

Kane, A. A., van Swol, L. M., & Sarmiento-Lawrence, I. G. (2023). Emotional contagion in online groups as a function of valence and status. Computers in Human Behavior, 139, 107543.

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as a museum guide–design and evaluation of a real-world application. In Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5 (pp. 329-343). Springer Berlin Heidelberg.

Korte, Russell. (2007). A review of social identity theory with implications for training and development. Journal of European Industrial Training. 31. 166-180.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision support systems, 48(2), 354-368.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of the speaker's voice. Journal of educational Psychology, 95(2), 419.

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

Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., & Frick, N. R. (2021). Understanding collaboration with virtual assistants–the role of social identity and the extended self. Business & Information Systems Engineering, 63, 21-37.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of social issues, 56(1), 81-103.

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. Telematics and Informatics, 35(1), 136-147.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in information retrieval, 2(1–2), 1-135.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. Computers in human behavior, 15(2), 123-142.

Pathak, S., Islam, S. A., Jiang, H., Xu, L., & Tomai, E. (2022). A survey on security analysis of Amazon echo devices. High-Confidence Computing, 100087.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. Cambridge, UK, 10, 236605.

Schweitzer, F., Belk, R., Jordan, W., & Ortner, M. (2019). Servant, friend, or master? The relationships users build with voice-controlled smart devices. Journal of Marketing Management, 35(7-8), 693-715.

Sikora, R. T., & Chauhan, K. (2012). Estimating sequential bias in online reviews: a Kalman filtering approach. Knowledge-Based Systems, 27, 314-321.

Silvarebels, C., & Ribeiro, B. (2003, July). The importance of stop word removal on recall values in text categorization. In Proceedings of the International Joint Conference on Neural Networks, 2003. (Vol. 3, pp. 1661-1666). IEEE.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311, 18-38.

Slota, S. C., Yi, S., Fleischmann, K. R., Bailey, J., & Watkins, S. C. (2021, November). Methods for Eliciting Feedback about AI and Racial Equity: How Black and Latinx Youth Interact with Digital Assistants. In TMS Proceedings 2021. PubPub.

Ulhøi, J. P., & Nørskov, S. (2022). The emergence of social robots: Adding physicality and agency to technology. Journal of Engineering and Technology Management, 65, 101703.

Wald, R., Piotrowski, J. T., Araujo, T., & van Oosten, J. M. (2023). Virtual assistants in the family home. Understanding parents' motivations to use virtual assistants with their Child (dren). Computers in Human Behavior, 139, 107526.

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding the bag-of-words model: a statistical framework.

Zhao, J., & Rau, P. L. P. (2020). Merging and synchronizing corporate and personal voice agents: Comparison of voice agents acting as a secretary and a housekeeper. Computers in Human Behavior, 108, 106334.

Appendix

```
#clean environment
rm(list=ls())

#load libraries
library(readxl)
library(dplyr)
library(stringr)
library(DT)
library(NLP)
library(tidytext)
library(dplyr)
library(stringr)
library(sentimentr)
library(ggplot2)
library(RColorBrewer)
library(readr)
library(SnowballC)
library(tm)
library(wordcloud)
library(reticulate)
library(crfsuite)
library(textdata)
library(kableExtra)
library(vader)
library(stargazer)
```

Dataset and Data Pre-processing Steps

```
#read data
dataset_reviews <- read_excel("Dataset - Nova IMS - Vitória Fontana -
m20200261.xlsx")

#create a new column that includes year of the review
dataset_reviews <- dataset_reviews %>%
  mutate(year = substr(date, nchar(date) - 3, nchar(date)))

#tokenizes words, removes stop words and any punctuation.
#and creates individual row for each word
words <- dataset_reviews %>%
  select(c("id", "profileName", "rating", "text", "year")) %>%
  unnest_tokens(word, text) %>%
  filter(!word %in% stopwords("en"), str_detect(word, "^[a-z']+$"))

#downlaods afinn library that includes sentiment score for each word
afinn <- get_sentiments("afinn")

#merges afinn library with review dataframe
reviews_afinn <- words %>%
```

```
  inner_join(afinn, by = "word")

#groups by id and profile name
afinn_score <- reviews_afinn %>%
  group_by(id, profileName) %>%
  summarise(mean_sent_score = mean(value))

#merge afinn score df with review df
merged_df <- dataset_reviews %>%
  inner_join(afinn_score, by = c("id", "profileName"))
```

Descriptives

```
#groups reviews by year and calculates the average rating per year
reviews_yearly <- merged_df %>%
  group_by(year) %>%
  summarize(Count = n(), average_rating = round(mean(rating),2))

colnames(reviews_yearly) <- c("Year", "Number of Reviews", "Average
Rating")

reviews_yearly_table <- reviews_yearly %>%
  kbl(align = c(rep("c",3))) %>%
  kable_classic(full_width = F, position = "center",  html_font =
"Cambira") %>%
  kable_styling(bootstrap_options = "striped") %>%
  column_spec(1, width = "1.5cm") %>%
  column_spec(2, width = "1.5cm") %>%
  column_spec(3, width = "1.5cm")

#groups reviews by rating and counts them
reviews_rating <- merged_df %>%
  group_by(rating) %>%
  summarize(Count = n())

rating_yearly_plot <-
  ggplot(data = reviews_rating, aes(x = rating, y = Count)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_classic()+
  theme(text = element_text(size = 14, color = "black"),
        axis.line = element_line(color = "black"),
        axis.text = element_text(size = 14, color = "black"),
        plot.title = element_text(size = 24, color = "black"),
        plot.subtitle = element_text(size = 18, color = "black"),
        plot.caption = element_text(size = 12, color = "black"),
        legend.title = element_text(size = 14, color = "black"),
        legend.text = element_text(size = 14, color = "black"),
        legend.key = element_rect(fill = "#F7F7F7"),
        panel.grid.major.y = element_line(color = "grey", size = 0.5),
        legend.position = "top", legend.box = "horizontal") +
  scale_y_continuous(limits = c(0,7000))+
```

```
  xlab("Rating") +
  ylab("Number of Reviews")
```

Most Frequent Positive and Negative Words

```
#groups by word, counts the number of occurrences and calculates the
average rating when the word is used
#filters good words with afinn sentiment score larger 0 and shows the
top 10 positive words
good <- reviews_afinn %>%
  group_by(word) %>%
  summarise(mean_rating = round(mean(rating),2), score = max(value),
count_word = n()) %>%
  filter(score >0) %>%
  arrange(desc(count_word))

good15 <- head(good, n = 15)

good15 <- cbind(1:15, good15)

colnames(good15) <- c("", "Word", "Average Rating", "Sentiment Score",
"Frequency")

good_table <- good15 %>%
  kbl(align = c(rep("c",5))) %>%
  kable_classic(full_width = F, position = "center",  html_font =
"Cambira") %>%
  kable_styling(bootstrap_options = "striped") %>%
  column_spec(1, width = "1.5cm") %>%
  column_spec(2, width = "2.5cm") %>%
  column_spec(3, width = "2.5cm") %>%
  column_spec(4, width = "2.5cm") %>%
  column_spec(5, width = "2.5cm")

wc_good_words <-
  wordcloud(good$word,
            freq = good$count_word, min.freq = 1, max.words = 50,
scale = c(2.2,1),
            colors=brewer.pal(8, "Dark2"), random.color = T,
random.order = F)
```

```
#groups by word, counts the number of occurrences and calculates the
average and rating when the word is used
#filters negative words with afinn sentiment score larger 0 and shows
the top 10 negative words
bad <- reviews_afinn %>%
  group_by(word) %>%
  summarise(mean_rating = round(mean(rating),2), score = max(value),
count_word = n()) %>%
  filter(score < 0) %>%
```

```
  arrange(desc(count_word))

bad15 <- head(bad, n = 15)

bad15 <- cbind(1:15, bad15)

colnames(bad15) <- c("", "Word", "Average Rating", "Sentiment Score",
"Frequency")

bad_table <- bad15 %>%
  kbl(align = c(rep("c",5))) %>%
  kable_classic(full_width = F, position = "center",  html_font =
"Cambira") %>%
  kable_styling(bootstrap_options = "striped") %>%
  column_spec(1, width = "1.5cm") %>%
  column_spec(2, width = "2.5cm") %>%
  column_spec(3, width = "2.5cm") %>%
  column_spec(4, width = "2.5cm") %>%
  column_spec(5, width = "2.5cm")

wc_bad_words <-
  wordcloud(bad$word,
            freq = bad$count_word, min.freq = 1, max.words = 50, scale
= c(2.2,1),
            colors=brewer.pal(8, "Dark2"), random.color = T,
random.order = F)
```

Distribution of Words based on Sentiment Score

```
word_summary <- reviews_afinn %>%
  group_by(word) %>%
  summarise(mean_rating = mean(rating), score = max(value), count_word
= n()) %>%
  arrange(desc(count_word))

all_words_afinn_dist <-
  word_summary %>%
  group_by(score) %>%
  summarise(count= sum(count_word))

all_words_afinn_plot <-
  ggplot(data = all_words_afinn_dist, aes(x = score, y = count)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_classic() +
  theme(
    text = element_text(size = 14, color = "black"),
    axis.line = element_line(color = "black"),
    axis.text = element_text(size = 14, color = "black"),
    plot.title = element_text(size = 24, color = "black"),
    plot.subtitle = element_text(size = 18, color = "black"),
    plot.caption = element_text(size = 12, color = "black"),
    legend.title = element_text(size = 14, color = "black"),
```

```
    legend.text = element_text(size = 14, color = "black"),
    legend.key = element_rect(fill = "#F7F7F7"),
    panel.grid.major.y = element_line(color = "grey", size = 0.5),
    legend.position = "top",
    legend.box = "horizontal"
  ) +
  scale_x_continuous(limits = c(-5, 5), breaks = seq(-5,5, by = 1)) +
  labs(x = "AFINN Sentiment Score" , y = "Number of Words")
```

Most Frequent Words in Positive and Negative Reviews

```
good_reviews <- reviews_afinn %>%
  filter(rating == 5)

good_reviews_ranked <-
  good_reviews %>%
  group_by(word) %>%
  summarise(mean_rating = round(mean(rating),2), score = max(value),
count_word = n()) %>%
  arrange(desc(count_word))

good_reviews_ranked <- head(good_reviews_ranked, n = 15)

good_reviews_ranked <- cbind(1:15, good_reviews_ranked)

colnames(good_reviews_ranked) <- c("", "Word", "Average Rating",
"Sentiment Score", "Frequency")

good_reviews_ranked_table <- good_reviews_ranked %>%
  kbl(align = c(rep("c",5))) %>%
  kable_classic(full_width = F, position = "center",  html_font =
"Cambira") %>%
  kable_styling(bootstrap_options = "striped") %>%
  column_spec(1, width = "1.5cm") %>%
  column_spec(2, width = "2.5cm") %>%
  column_spec(3, width = "2.5cm") %>%
  column_spec(4, width = "2.5cm") %>%
  column_spec(5, width = "2.5cm")

bad_reviews <- reviews_afinn %>%
  filter(rating %in% c(1, 2))

bad_reviews_ranked <-
  bad_reviews %>%
  group_by(word) %>%
  summarise(mean_rating = round(mean(rating),2), score = max(value),
count_word = n()) %>%
  arrange(desc(count_word))

bad_reviews_ranked <- head(bad_reviews_ranked, n = 15)

bad_reviews_ranked <- cbind(1:15, bad_reviews_ranked)
```

```
colnames(bad_reviews_ranked) <- c("", "Word", "Average Rating",
"Sentiment Score", "Frequency")

bad_reviews_ranked_table <- bad_reviews_ranked %>%
  kbl(align = c(rep("c",5))) %>%
  kable_classic(full_width = F, position = "center",  html_font =
"Cambira") %>%
  kable_styling(bootstrap_options = "striped") %>%
  column_spec(1, width = "1.5cm") %>%
  column_spec(2, width = "2.5cm") %>%
  column_spec(3, width = "2.5cm") %>%
  column_spec(4, width = "2.5cm") %>%
  column_spec(5, width = "2.5cm")
```

Boxplot and Regression Analysis

```
correlation_rat_sent <- cor(merged_df$rating,
merged_df$mean_sent_score)

merged_df$rating_group <- cut(merged_df$rating, breaks = c(0, 1, 2, 3,
4, 5),
                              labels = c("1 Star", "2 Star", "3
Star", "4 Star", "5 Star"))

meant_sent_boxplot <-
  ggplot(merged_df, aes(x = mean_sent_score, y = rating_group)) +
    geom_boxplot() +
    labs(x = "Mean Sentiment Score", y = "Rating") +
    theme_minimal()+
    theme(text = element_text(size = 14, color = "black"),
          axis.line = element_line(color = "black"),
          axis.text = element_text(size = 14, color = "black"),
          plot.title = element_text(size = 24, color = "black"),
          plot.subtitle = element_text(size = 18, color = "black"),
          plot.caption = element_text(size = 12, color = "black"),
          legend.title = element_text(size = 14, color = "black"),
          legend.text = element_text(size = 14, color = "black"),
          legend.key = element_rect(fill = "#F7F7F7"),
          panel.grid.major.y = element_line(color = "grey", size =
0.5),
          legend.position = "top", legend.box = "horizontal")+
    scale_x_continuous(limits = c(-5, 5), breaks = seq(-5,5, by = 1))


lin_model <- lm(rating ~ mean_sent_score, data = merged_df)

reg_table <-
  stargazer(
    lin_model,
    type = "latex",
    title = "Regression Results",
```

```
    style = "aer",
    header = FALSE,
    dep.var.labels = c("Rating"),
    covariate.labels = c("Mean Sentiment Score"),
    no.space=TRUE,
    out = "regression.html"
  )
```

NRC Analysis

```
nrc <- get_sentiments("nrc")

sentiments <- words %>%
  inner_join(nrc, "word") %>%
  count(word, sentiment, sort=TRUE)

nrc_sent_plot <-
  ggplot(data=sentiments, aes(x=reorder(sentiment, -n, sum), y=n)) +
    geom_bar(stat = "identity", fill = "steelblue") +
    theme_classic()+
    labs(x="Sentiment", y="Frequency") +
    theme(
      text = element_text(size = 14, color = "black"),
      axis.line = element_line(color = "black"),
      axis.text = element_text(size = 14, color = "black"),
      plot.title = element_text(size = 24, color = "black"),
      plot.subtitle = element_text(size = 18, color = "black"),
      plot.caption = element_text(size = 12, color = "black"),
      legend.title = element_text(size = 14, color = "black"),
      legend.text = element_text(size = 14, color = "black"),
      legend.key = element_rect(fill = "#F7F7F7"),
      panel.grid.major.y = element_line(color = "grey", size = 0.5),
      legend.position = "top",
      legend.box = "horizontal"
    )
```