“PREDICTING THE IMPACTS OF BUSINESS ON THE SOCIAL MEDIA”

Student Name	**Niharika Ganji**
Student ID	**32032536**
Supervisor Name	**Aitsam Muhammad**

Abstract

Sentiment analysis is a method of determining the emotional state expressed in information or text by categorising it into three categories: positive, negative, and neutral. The emergence of social media has increased the amount of sentiment data available on the internet. People on various platforms have the opportunity to express their opinions on a variety of products, services, and other topics on social media. This has led to the emergence of sentiment analysis as one of the most interesting areas of research in natural language processing (NLP). The proposed model was shown for sentiment analysis via text blob and pattern. The five small business-oriented companies' reviews are analysed and compared to determine which one is most satisfactory based on the sentiment score ratio between the total number of web-scraped sentences and positive words.

Acknowledgement

I would like to express my gratitude for the support and inspiration provided by the academics at Sheffield Hallam University, especially Aitsam Muhammad. The students who volunteered to take the surveys and participate in the interviews were essential to the success of this study. Additionally, friends and family contributed to the project's success by continuing to support it and by editing the work as it was being completed.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background of the study

The landscape of business and consumer interactions has undergone a significant transformation in the modern digital era due to social media's pervasive influence. An unprecedented reliance on social media platforms for marketing, communication, and brand promotion has resulted in businesses of all sizes quickly expanding into the digital space. This change has been brought about by the realisation that social media provides a direct channel to a huge and diverse audience, enabling businesses to engage with customers on an unheard-of scale (Nemes and Kiss, 2021). The use of artificial intelligence (AI) and machine learning (ML) techniques, such as sentiment analysis, has emerged as a crucial tool in this context for comprehending and foreseeing the effects of business activities on social media.

In recent years, sentiment analysis has gained popularity among organizations, companies, and governments in addition to scholars. The internet has become the major source of information for the general population as a result of its rising popularity. Users use a variety of internet sites to share their ideas and viewpoints. Platforms for social interaction, opinion sharing, and sentiment expression have become essential elements of people's daily lives. Businesses now have a rare chance to interact directly with their target market thanks to this phenomenon, but there are drawbacks as well. The sheer volume of user-generated content on social media platforms calls for creative methods to glean insightful information. Here's where AI and ML come into play, completely changing how companies use these platforms' power (Aljedaani et al. 2022).

The main goal of this study is to use sentiment analysis to explore the complex interactions that exist between businesses and social media. This method of analysis makes use of AI and ML to evaluate the feelings, attitudes, and opinions contained within textual data, providing a window into the general sentiment of consumers (Chauhan et al., 2021). Businesses can better understand how the public views their actions, goods, and services by analysing sentiment.

The enhancement of small business-oriented companies has been designated as the focal point of this study. Enterprises like "Bettys," "The Yorkshire Soap Company," "Black Yak," and "Masons Yorkshire Gin" epitomize the vitality of small businesses that have harnessed the potential of social media to propel their expansion. These firms often leverage customer feedback to refine their offerings and enhance overall customer satisfaction. Nevertheless, the manual collection and analysis of this feedback can prove to be arduous and time-intensive. In this context, the utilization of AI-driven sentiment analysis emerges as a transformative solution with the potential to revolutionize the process.

Sentiment analysis is the process of gathering and examining people's thoughts, beliefs, and perceptions on a range of issues, products, and services. The volume of comments and evaluations about routine activities has dramatically increased on social media platforms, blogs, and other Internet-based apps. For businesses, governments, or individuals to acquire information and make wise judgments, the views of people can be valuable. However, there are plenty of challenges with sentiment analysis and the rating process. These challenges make it difficult to accurately evaluate sentiment and establish the proper sentiment polarity (positive or negative). Sentiment analysis uses text mining and natural language processing (NLP) to extract and evaluate subjective information from text (Nemes and Kiss, 2021).

As social media sites have grown in popularity, several other disciplines have emerged that specialize in analyzing them and their content to obtain the data they want. Sentiment analysis focuses on determining readers' attitudes toward a text by examining its content. Since it falls under the umbrella of Natural Language Processing (NLP), a lot of early research must have been done in this area. However, it is still under development and won't be completed before the turn of the millennium. There are a number of difficulties in sentiment analysis, such as the informal writing style of people, the usage of sarcasm, and the use of irony. The fact that there are several terms in different languages with different meanings and orientations depending on the context and domains they are employed in only makes these difficulties worse. There aren't enough tools and resources available for all languages as a result. Simplicity and sarcastically in writing have received a lot of attention recently from thinkers. This essay will give an overview of the numerous

problems that sentiment analysis faces, along with the solutions, tools, and algorithms that are employed (Basarslan and Kayaalp, 2020).

It appears that several of the methods utilized in sentiment analysis, such as machine learning and transformer learning, are not included in current surveys. This article addressed every aspect of the assignment, but it stands out from other studies since it includes the most often used methodologies. Other studies focus solely on sentiment analysis related to a task, various difficulties, or a particular concern like product evaluations. This study examines sentiment analysis from a variety of angles since it covers a wide range of relevant research topics, including issues, applications, tools, and methodologies. It's fantastic for experts and novices alike because anyone can learn a lot about it in one paper.

1.2 Problem Statement

The profound influence of social media on the rapidly changing business landscape has completely changed how businesses interact with their target markets and promote their goods and services. Businesses face the difficult challenge of comprehending and anticipating the complex effects of their activities in this digital space as they increasingly recognise the potential of social media platforms as tools for brand exposure, customer interaction, and growth (Piedrahita-Valdes et al. 2021). Deciphering the complex relationships between businesses' social media activities and the opinions of the customers they serve is the main issue this study attempts to solve.

The urgent issue in this situation is best stated as follows: How can companies use social media platforms to effectively promote their products while also understanding the feelings, viewpoints, and reactions of their customer base? For small business-oriented organisations like "Bettys," "The Yorkshire Soap Company," "Black Yak," and "Masons Yorkshire Gin," where growth and success heavily depend on forging a resonance with their clientele in the virtual space, this challenge becomes particularly pertinent.

Due to the overwhelming volume of unstructured textual data generated by the social media industry's explosive growth, it is now nearly impossible for businesses to manually process

and thoroughly analyse consumer sentiment. Furthermore, it can be difficult to distinguish between flattering compliments and sincere expressions of satisfaction. In order to overcome these difficulties, this research employs AI and ML techniques, particularly sentiment analysis, as a means of extracting insightful data from the substantial digital footprint that users leave across various online platforms (Aljedaani et al. 2022).

The choice of appropriate sentiment analysis tools and methodologies is also a problem. Finding the most precise and trustworthy options that fit the needs of small businesses can be difficult given the variety of tools at hand. Making educated choices about the tools that will accurately interpret and categorise the range of emotions expressed in customer reviews, comments, and interactions is the difficult part.

Additionally, there are dangers introduced by sentiment analysis's inherent subjectivity. Complex expressions must be labelled as positive, negative, or neutral with a nuanced understanding of language and context (Asif et al. 2020). Misinterpretations or incorrect classifications may produce flawed insights that misdirect business strategies, resulting in opportunities lost or efforts put in the wrong direction.

In order to accurately predict and understand the effects of businesses on social media, a comprehensive approach that combines AI, ML, and sentiment analysis is required. By enabling small business-oriented companies to make data-driven decisions that are in line with the preferences and expectations of their online audience, this research aims to close the gap between businesses' digital initiatives and consumer sentiments.

1.3 Research aims and Objectives

Aim:

The purpose of this project is to look into the potential effects of social media on commercial enterprises, especially those geared towards small businesses like Betty's, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin. The research will make use of machine learning (ML) and artificial intelligence (AI) methodologies to accomplish this.

Objectives:

- To develop a methodical approach to web content extraction from social websites, such as blogging and review platforms, using web scraping techniques to ensure the collection of pertinent data for analysis.
- To create a thorough framework for data processing that uses regular expression techniques to efficiently recognise and separate meaningful sentences from the extracted content, Investigate the use of Python AI scripts to simplify tokenization, sentence cleaning, and other crucial preprocessing steps.
- To analyse and interpret the sentiments embedded in the textual data, use established Python sentiment analysis tools like Pattern and TextBlob (Aljedaani et al. 2022). Examine how these tools can be used to correctly categorise feelings as positive, negative, or neutral.
- To consider how the results of sentiment analysis can be used in practice to raise client satisfaction with a business's operations. Examine the ways in which sentiment insights can aid in the development of customer-centric strategies and serve as a strategic decision-making tool.

1.4 Research Questions

1. How does the web scraper extract the web contents from social websites like review websites or some blogging-type websites?
2. How the extracted contents are processed by the method of regular expression for picking up idle sentences without any emoji or any symbolic expressions and how the Python AI scratch is used to serve as a tool for implementing sentence cleaning, tokenization, etc.
3. How are the sentiments analysed through the pattern and TextBlob, which are mostly used as Python sentiment tools?
4. How are sentiments helping to improve customer satisfaction with the company's activities?

1.5 Research rationale

Social media's pervasive influence in modern society has fundamentally changed how companies interact with their clients, market their goods, and build their brand identities. Businesses need to successfully navigate this complex environment as they come to recognise social media's potential as a powerful engagement tool (Basarslan and Kayaalp, 2020). In this context, the fusion of sentiment analysis and artificial intelligence (AI) presents a novel way to not only understand but also anticipate the subtle effects of business activities on social media platforms. This research justification explains the critical importance of examining social media's effects on businesses and highlights the crucial role AI-driven sentiment analysis plays in fostering data-driven strategies and raising customer satisfaction.

Recognising the Situation

Social media platforms have moved beyond their function as simple channels for communication and transformed into vibrant hubs where companies interact with their customers on a scale that was previously unattainable. User-generated content is abundant in the digital world and captures opinions, feelings, and sentiments. Businesses can use this information to improve their strategies, offerings, and customer experiences (Vashishtha and Susan, 2019). But to extract meaningful insights from the vast amounts of unstructured data, sophisticated methodologies are required. When AI and sentiment analysis work together, a useful solution can be found.

Using AI and sentiment analysis to close the gap

The key to this study is bridging the gap between businesses' social media initiatives and consumer sentiments by using AI and sentiment analysis. AI improves human abilities by analysing large datasets, identifying patterns, and producing useful insights (Drus and Khalid, 2019). Additionally, a deep understanding of how customers view, respond to, and perceive business activities is provided by sentiment analysis, a subfield of artificial intelligence (AI). Sentiment analysis enables the identification and classification of sentiments within textual data.

The Need for Little Businesses:

The analysis of effects is especially important for small businesses because they frequently have limited resources and rely heavily on customer loyalty. Companies like Bettys, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin are the best examples of this demographic, and it is strategically important to comprehend how they interact online (Khan et al. 2020). These companies can improve their marketing tactics, boost customer satisfaction, and spur growth by revealing sentiment patterns in their interactions.

Research Justification:

1. **Making Strategic Decisions:** The foundation of the research's justification is the idea that a well-informed strategy complements consumer sentiment. Businesses can make decisions that resonate with their target audience by deciphering the undercurrents of sentiment in consumer discussions.
2. **Enhanced Customer Satisfaction:** By exposing the feelings that surround business activities, one can better predict what customers will expect. This encourages brand loyalty and repeat business, which are crucial for small businesses looking to grow sustainably.
3. **Data-Driven Insights:** By combining sentiment analysis and AI, data-driven insights are provided to businesses. These insights go beyond intuition and speculative opinion, empowering businesses to make wise decisions and stay clear of potential pitfalls.
4. **Competitive Advantage:** Companies that use sentiment analysis have an advantage over their competitors. They are able to foresee changes in consumer preferences and adjust their offerings accordingly, outperforming their rivals.

1.6 Significance of the study

The proposed study, which examines how social media affects businesses through the lenses of sentiment analysis and artificial intelligence, is extremely important in the context of modern business. The results of this study have the potential to have a significant impact

on strategies, choices, and customer satisfaction as businesses around the world navigate the digital space to connect with their target audience. The importance of the study stems from its potential to transform business practices, strengthen small businesses, and advance the discussion on the symbiotic relationship between technology and successful business outcomes (Krishen et al. 2021).

- **Making Knowledgeable Decisions:** The study's investigation of AI and sentiment analysis equips businesses with data-driven insights that go beyond firsthand knowledge. Making informed decisions is facilitated by the capacity to directly infer consumer sentiments and opinions from digital interactions. Businesses can tailor marketing campaigns, product offerings, and engagement strategies to match the expectations and preferences of their target market once they have access to these insights.
- **Enhanced Customer Satisfaction:** This study's potential to increase customer satisfaction is one of its most important contributions. Businesses can proactively address concerns, adjust to changing demands, and enhance overall customer experiences by analysing sentiment patterns in consumer conversations (Derakhshan and Beizy, 2019). In turn, this fosters brand loyalty and strengthens connections, two essential elements for long-term success.
- **Empowering Small Businesses:** Small businesses face particular difficulties that are frequently hampered by visibility and resources. The significance of the study is increased by the emphasis on organisations like Bettys, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin. These businesses can use sentiment analysis to sharpen their strategies, gain a competitive edge, and strategically place themselves in the digital landscape—possibly levelling the playing field with more established competitors.
- **Support for academia:** The study is in line with academic efforts to comprehend how business dynamics and technology interact. Its empirical findings, methodologies, and conceptual framework add to the body of knowledge that is constantly expanding on how sentiment analysis and AI are used to create contemporary business strategies. These insights can be used by academia to inform

curricula, stimulate additional study, and encourage critical debates on the implications of technology-driven transformations.

- **Business Future Research:** AI and sentiment analysis will inevitably be integrated into business procedures as they develop. The study's conclusions provide a glimpse of how business research will develop in the future, when the fusion of technology and consumer insights will become essential. The findings of the study can serve as a roadmap for academics, professionals, and decision-makers as they explore new facets of marketing trends, consumer behaviour, and business strategy.

1.7 Structure of the study

In the introduction of the study, the significance of social media's influence on businesses is set in context, as is the rise of AI and sentiment analysis as tools for predicting and comprehending consumer sentiment. The research's focal point and objectives are briefly described, along with the title and goals of the study.

- The literature review synthesises existing research on the relationship between social media, artificial intelligence (AI), and business outcomes in a critical manner. It highlights the importance of sentiment analysis in understanding consumer attitudes and behaviours. The review highlights the study's relevance by focusing on small businesses and their potential for expansion using sentiment-driven tactics.
- **Research Methodology:** This section describes the quantitative research strategy used, which includes web scraping to collect data from various websites. The use of regular expression techniques for data processing is discussed, as is the use of AI-driven sentiment analysis tools. Procedures for gathering data and ethical issues are also covered.
- **Data Analysis and Findings:** The data analysis shows patterns of sentiment in consumer interactions and insights from sentiment analysis. Visual aids improve the way that results are presented while providing a thorough understanding of consumer attitudes towards commercial activities.

- **Conclusion and Suggestions:** The conclusion highlights the transformative potential of sentiment analysis in business decision-making by summarising key findings. Underscoring the usefulness of the study's findings, recommendations are given for small businesses to use sentiment insights for customer engagement and growth.

Chapter 2: Literature Review

2.1 Overview of the chapter

The literature review chapter discusses an in-depth analysis of the increasing impact of social media in the business environment in this modern age. It emphasises how social media marketing has developed from an untraditional strategy to a useful tool, enabling businesses to combine with an expansive global audience and improve brand distinction and sales. The review highlights the modification of communication between industries and customers, leading to improved customer attention and the possibility of acquiring useful insight into customer conduct. In today's dynamic market, harnessing the probable impact of social media is important for businesses to stay competitive and relevant. The chapter seamlessly transitions to the study methodology and deliverables that will investigate social media's impact on key performance indicators, future customer trends, and the significance of emotion analysis and predictive awards.

2.2 Introduction to Social Media's Impact on Business and its promotion

In the modern digital era, social media has revolutionised the way businesses operate. Through technology, the world becomes increasingly interconnected, and businesses are to expand their reach by leveraging the power of social media, engage with customers, and stay competitive (Aichner et al., 2021). This literature review surveys the growing role of industries in social media, its importance in the traditional business topography, and the ideals of the research about the impact of social media on businesses.

The Increasing Role of Business in Social Media: In the past few years, there has been an important shift in how businesses perceive and utilise social media platforms, including how and from which perspective. From earlier suspicion to now being an important aspect of commerce and consumer arrangement techniques, social media has evolved into a strong mechanism for businesses across different enterprises. Businesses have identified the possibility of social media reaching extensive and various audiences, delivering a forum for immediate communications with their target market (Krishen et al. 2021).

The importance of social media in the present business landscape

The importance of social media in the present business landscape cannot be overemphasized. With over 4.5 billion dynamic social media users globally, these media have evolved into a virtual marketplace, promoting businesses to engage with customers on an individual level and form brand beliefs. Social media performs as a knowledge hub where customers pursue reviews and outcome suggestions and link with like-minded people. Businesses that virtually leverage social media achieve helpful insight into consumer choices, manners, and market trends, helping them create data-driven conclusions and stay relevant in a continuously growing market. Here are some significant examples:

1. **Nike:** Nike, the sportswear giant, has a strong presence on social media, specifically on media like Instagram and Twitter (Yasa et al. 2020). They utilise social media to showcase their trendy products, cooperate with fans, and engage with their huge international fan base. Nike's passionate and visually attractive content resonates with its target audience, strengthening its trademark image.
2. **Starbucks:** Starbucks is known for its innovative and fascinating social media movements. They use media like Instagram and Facebook to share attractive pictures of their coffee and other offerings, run interactive competitions, and encourage seasonal products. Starbucks' social media existence recreates an influential role in supporting consumer belief and moving foot gridlock to their marts. For their innovative products, they are well known in business marketing.
3. **Wendy's:** Wendy's, the fast-food chain, is well-known for its humorous and funny interactions on Twitter. They employ an active repartee with consumers and even rival brands, acquiring prominence for their sassy and amusing social media existence. Wendy's social media strategy has aided them in achieving a trustworthy following and supplementing trademark visibility (Chatterjee et al. 2020). Their product has become famous worldwide.
4. **Airbnb:** Airbnb operates social media to showcase unique and gorgeous travel goals, bits of knowledge, and concessions. They feature a user-generated range and engage with their society, creating a sense of belonging and faith.

Airbnb's social media approach highlights forming genuine and influential content to promote travel and investigation. Their fascinating travel destination is trustworthy, and therefore they gain faith from all the consumers.

5. **Coca-Cola:** Coca-Cola utilises social media to strengthen its international commerce movements and share news of positivity and enjoyment. They usually build interactive content and hashtags to contend with their audience and urn seasonal buildups and giveaways (Syaifullah et al., 2021). This brand is famous for its soft drinks and soda items.

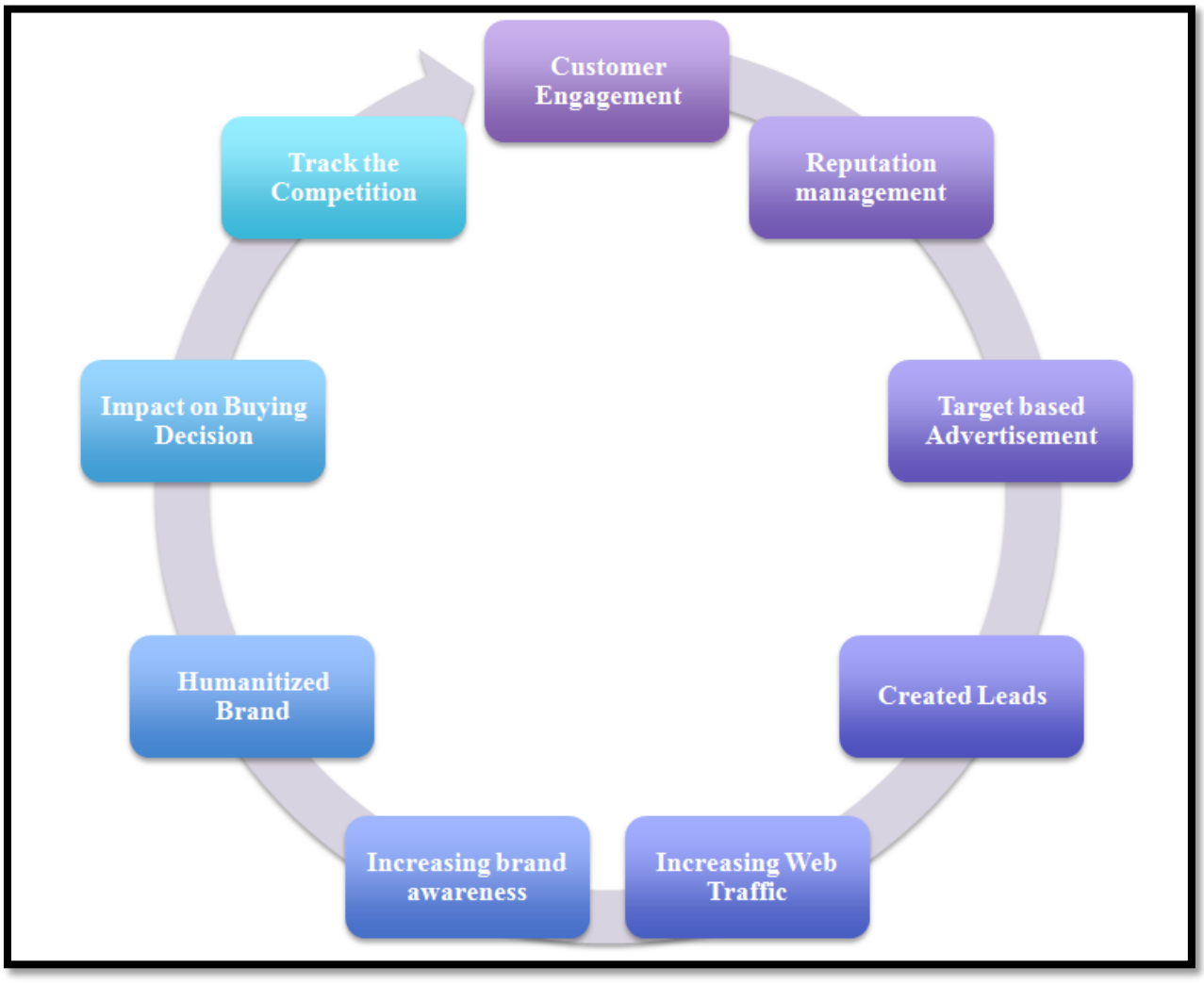


Figure 1: Advantages of social media for Businesses

(Source: Aichner et al. 2021, P-85)

2.3 Social media as a Marketing Tool for Businesses

Social media has converted the marketing topography, delivering businesses a strong platform to link with their target market, construct brand cognition, and drive sales. This literature review analyses the usage of social media for marketing and advertising, how businesses contend with their target market on media like Facebook, Twitter, and Instagram, and the importance of the overall utilization of social media for businesses (Talahaturuson et al. 2022, September).

Utilization of social media for Marketing and Promotion: social media has become an essential part of marketing techniques for businesses across different enterprises. Businesses use social media platforms to communicate promotional scope, commit to consumers, and promote trademark faithfulness (Aichner et al.2021). With the ability to build and communicate different content, such as posts, videos, stories, and live streams, businesses can share their brand messages in a vibrant and interactive technique.

Social media marketing allows businesses to get an expansive and varied audience, outperforming geographical limitations. This broad reach permits corporations to target distinct demographics, tailoring content to correspond to the attractions and importance of their conscious audience. Besides, the real-time essence of social media permits businesses to respond very fast to consumer feedback and questionings, improving customer fulfilment and brand prominence. It helps the business owner to achieve their goal.

Businesses' Use of Platforms like Trustpilot and other public review websites like Review.io: Trustpilot and are among the most famous public media platforms utilized by businesses for commerce and advertisement. By using these public media platforms businesses can grow well.

1. The Trustpilot platform is an online rating system designed to assist customers in making informed purchasing decisions and businesses in improving customer experience. Companies with Trustpilot profiles are assigned a Trust

Score ranging from one to five, based on past customer ratings. Consumers can also create a review that outlines their overall impression of the company. Established in 2007, Trustpilot, a Danish company, has over 120 million reviews on its platform, covering 529,000 companies as of December 31, 2020, and has raised the most capital in a London Stock Exchange IPO in almost a decade (Littlechild, 2021).

2. Millions of people around the world trust reviews from REVIEWS.io. Here the collection of review is independent, impartial, and honest. Reviews are important for businesses and useful for consumers. Businesses can ask for reviews from their real customers, and customers can easily find and read them on our public review pages.

Statistics on the Percentage of Adults Using Social Media and Its Implications for Businesses According to a Pew Research Centre survey conducted in 2021, around 72% of adults in the United States operate social media (Pew Research Centre, 2021). This statistic highlights the huge potential social media has for businesses to confront their target market. The overall use of social media among adults suggests that businesses can connect with a considerable portion of their consumer base through these media. Social media supplies businesses with beneficial information and an understanding of customer manners, preferences, and tendencies, permitting them to purify their trade methods and tailor offerings to satisfy customer requirements. Likewise, the increasing dependence on social media as a reference for data and suggestions emphasises the significance of businesses designing a powerful and reasonable online platform. Favourable reviews and word-of-mouth acceptance on social media can seriously impact consumers' purchasing judgements, compelling brand belief, and advocacy (Ye et al. 2021). Besides, the real-time essence of social media permits businesses to respond very quickly to consumer feedback and questions, improving customer fulfilment and brand prominence. It helps the business owner achieve their goal.

2.4 The Evolution of Communication Between Business and Consumers

The beginning of the internet, specifically social media, has revolutionised the way businesses and customers interact and communicate. This literature review analyses the development of communication between enterprises and customers due to the Internet, emphasising the role of social media. It also emphasises the remarkable features of the Internet that promote seamless communication and examines the benefits of this transformation for businesses with an international customer base.

Evolution of Communication between Businesses and Consumers Due to the Internet, especially social media: The Internet has given a new shape to the communication landscape, allowing businesses to employ customers directly and continuously. With the upgrade of social media platforms such as Facebook, Twitter, Instagram, and LinkedIn, the manner in which businesses communicate with their target audience has experienced a notable modification. Unlike conventional advertisement techniques, social media permits two-way communications, entitling businesses to specify a more accurate and personalised association with customers (Talahaturuson et al. 2022, September).

Social media has evolved into a virtual marketplace where businesses can share content, product updates, and advancements while also obtaining instant feedback from customers (Sima et al. 2020). This real-time communication encourages businesses to manage customer questions, references, and feedback promptly, enabling faith and belief.

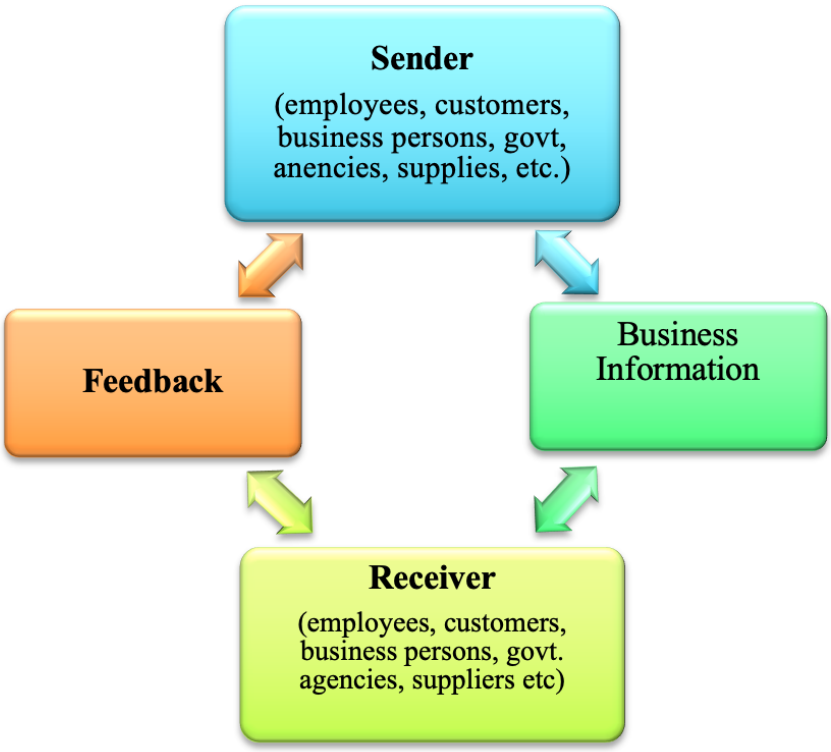


Figure 2: Communication between business and customers

(Source: Gupta et al.2021, P-145)

Distinctive Features of the Internet Facilitating Communication: several distinguishing characteristics of the Internet contribute to the development of contact between businesses and customers:

1. **Accessibility of Information:** The Internet delivers easy and fast access to expansive amounts of data. Customers can explore products, read reviews, and compare costs before making purchasing decisions (Gupta et al., 2021). Businesses can leverage this component by delivering detailed product information and transparent consumer reviews, building credibility and impacting customer choices.
2. **Global Reach:** The internet damages geographical borders, permitting companies to link with customers across the world. Social media platforms, individually, encourage arrangements with various audiences, allowing companies to tailor content to regional preferences and cultures.

3. **Real-time Communication:** Unlike traditional media, the Internet permits instant communication between businesses and customers. This real-time feature certifies businesses to manage customer questions promptly, address problems virtually, and support an active and amusing online presence.
4. **Interactivity and User-Generated Content:** The internet enables interactivity, allowing customers to actively contend with businesses via comments, likes, shares, and user-generated content. According to Gil-Gomez et al. (2020), user-generated content, in the form of studies, testimonials, and social media posts, improves brand credibility and affects purchasing decisions.

The evolution of communication between businesses and consumers emphasises the unique attributes of the Internet that encourage seamless communication and the benefits of this transformation. This literature review analyses the development of communication between enterprises and customers due to the Internet, emphasising the role of social media (Gil-Gomez et al. 2019). It also emphasises the remarkable features of the Internet that promote seamless communication and examines the benefits of this transformation for businesses with an international customer base.

Benefits of the Internet in terms of connecting businesses with an international customer base

The Internet proposes multiple benefits for businesses striving to connect with an international base of customers:

1. **Consumer insights and market research:** The Internet gives businesses access to comprehensive information on trends, behaviours, and preferences of consumers (Mills and Hair 2021). Via social listening tools and data analytics, businesses could achieve important understandings to inform their marketing strategies.
2. **Cost-efficient marketing:** Online marketing and communication are frequently more cost-efficient than contemporary methods of advertising. Platforms on social media enable businesses to connect with a wider audience with their targeted content, decreasing marketing costs.

3. **Expanded market reach:** The international reach of the Internet allows businesses to expand their market far away from local boundaries. Organisations could access new demographics and markets, directing growth and enhancing their share of the market.

4. **Improved consumer engagement:** The interactive characteristic of the Internet allows businesses to engage with customers in a significant manner. Interactive content, personalised responses, and authentic interactions promote a robust relationship with consumers.

2.5 Possible Influence of Social Media Marketing Tactics on Businesses

Social media has evolved as a strong tool for organisations, giving an exceptional opportunity to improve the visibility of the brand and achieve important understanding and engagement with consumers' behaviour (Coco and Eckert 2020). This review of the literature investigates the possible effect of social media on companies, concentrating on four key aspects: customer interaction, increased exposure, economical marketing, and insights into customer behaviour.

- **Customer Interaction:** Platforms on social media allow direct and real-time communication between customers and businesses. Consumers could express their concerns, inquiries, and feedback. On the other hand, businesses could answer promptly, promoting customer-centricity and a sense of clarity. Engagement in meaningful interactions with consumers assists in developing loyalty and trust. Organisations could use platforms on social media to give personalised service to their customers, which could have a deep effect on the perception of the brand and customer retention. Immediate responses to inquiries and determining problems publicly showcase a devotion to customer satisfaction, reinforcing the brand’s relationship with customers.

- **Increased Exposure:** Platforms on social media suggest businesses access a diverse and broad audience on an international scale. With billions of users who are active on different platforms, companies could reach the demographic they might

have floundered to engage with via contemporary channels of marketing. Social media enables businesses to transfer content, updates, and promotions, intensifying the message of the brand to a wider audience (Lawes 2023). Therefore, with the power of social sharing, companies could grow their reach. When consumers engage with the content of the brand and transfer it within their networks, it could generate a ripple impact, fostering organic development and oral marketing.

- **Economical Marketing:** In comparison with the channels of traditional marketing, social media proposes a means of promoting the brand cost-effectively. Making and sharing content on social media is frequently more cost-friendly than traditional advertising methods such as print ads or TV ads. However, advertising on social media platforms gives an option for advanced targeting, enabling businesses to connect with particular demographics, behaviours, and interests. This targeting ability makes sure that efforts in marketing are directed at the most appropriate audience, enhancing the investment return.

- **Insights into Customer Behaviour:** Platforms of social media produce mass data on the preferences, behaviour, and trends of the customer. This data about customers could be leveraged by businesses to achieve a more precise understanding of their target audience. Assessing comments, engagement patterns, and interactions on social media could give important feedback on the preferences and sentiments of the customers. Businesses could refine their offerings of products, evolve targeted campaigns, and tailor marketing messages by understanding the behaviours of consumers. Analysis of social media information enables businesses to detect developing trends and respond rapidly to changing markets, providing a competitive benefit.

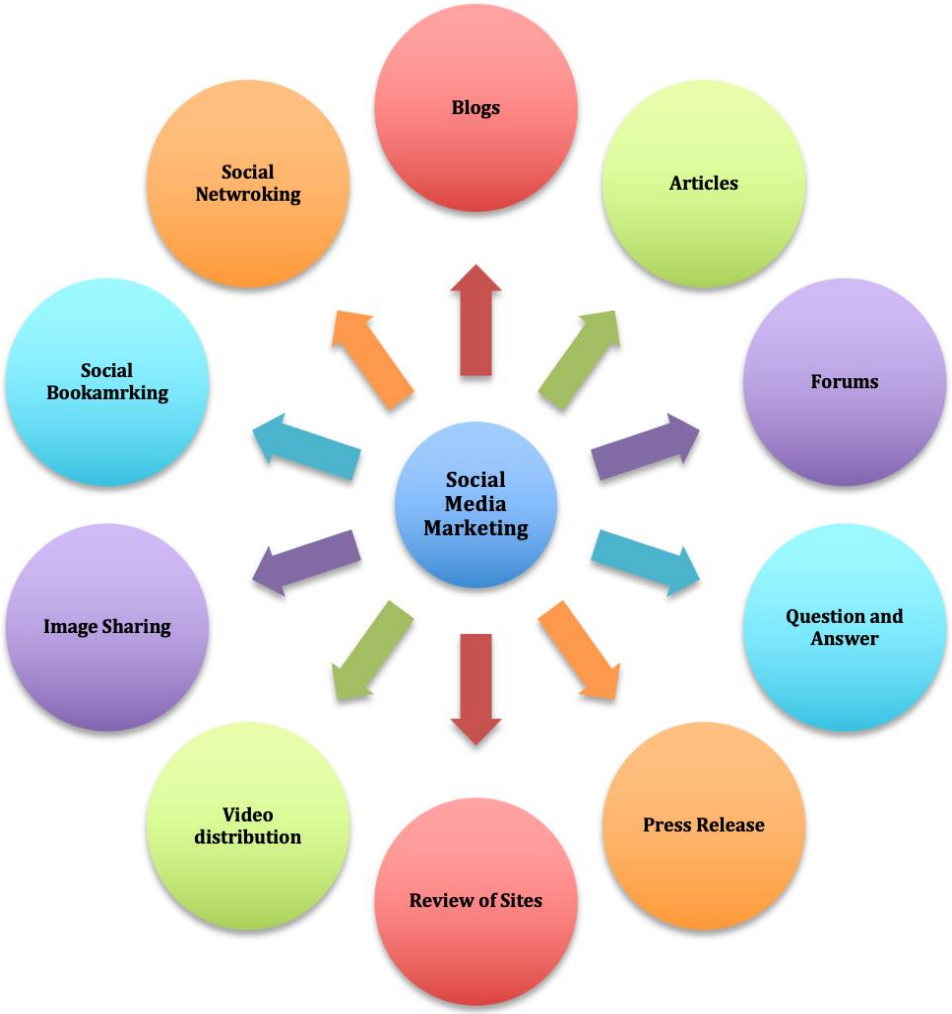


Figure 3: Types of Social Media Marketing

(Source: Gupta et al. 2021, P-68)

2.6 Future Implications of Social Media for Businesses

As social media is constantly growing, its effects on businesses are anticipated to undergo significant transformations in the future. This review of literature assesses the possible implications of social media on companies in the upcoming years, concentrating on enhanced competition, the requirement for creative and strategic utilisation, challenges and opportunities, and foreseeing the usage and perceptions of consumers.

Enhanced competition, The widespread adaptation of social media by companies is likely to enhance competition within these platforms. As more organisations determine the possibility of social media for customer engagement and brand promotion, the digital landscape will become more packed. As per Oliveira et al. (2020), businesses will have to compete for the attention of customers, making it important for brands to differentiate themselves from other competitive brands. Forecasting the possible rise in competition among businesses on platforms of social media depicts that advertising spending on social media is on the rise within organizations. This trend demonstrates that more brands are allocating their resources to marketing via social media, generating stronger competition for market share and audience engagement.

In an increasingly competitive environment of social media, companies need to utilise these platforms creatively and strategically to stand out. Businesses need to adapt innovative and unique methods to retain and capture the attention of their audiences, as having a simple presence on social media is not sufficient. Brands that leverage personalised content, creative campaigns, and data analytics will have a higher probability of cutting through the noise with their targeted market (Hu and Chaudhry 2020). Creating transparent and authentic connections with customers via platforms like social media will be crucial for developing brand loyalty.

Obstacles and opportunities

The future of social media suggests a multitude of opportunities and challenges for businesses.

Assess the possible opportunities:

1. **International reach:** Platforms on social media outperform geographical limitations, offering organisations an opportunity to engage and reach an international base of customers. According to Moriuchi (2019), enlarging new demographics and markets becomes more attainable via strategic utilisation of social media.

2. **Improved customer insights:** The data analytics abilities of social media will enable brands to achieve a more precise understanding of the behaviours, preferences, and trends of customers. This data could disclose the marketing techniques, improvement of the customer experience, and development of the product.

3. **Influencer marketing:** Collaborating with influencers on social media could strengthen the message of the brand and develop its reach. The marketing of influencers will present possibilities for brands to leverage collaboration with influencers efficiently.

Evaluate the possible challenges.

1. **Managing online reputation:** The virility and speed of social media could rapidly heighten the crisis of the brand and negative publicity (Meire et al. 2019). Organisations will require efficient strategies for crisis management to safeguard their reputation in the face of challenges that occur online.

2. **Security and privacy concerns:** As usage of social media increases, security and privacy concerns might occur, affecting the trust of consumers in businesses. Brands need to be proactive in terms of addressing privacy challenges and safeguarding consumer information.

Expecting consumer usage and perceptions

As platforms for social media constantly develop, usage patterns and perceptions of customers will form approaches and strategies for businesses. Customer behaviour on platforms of social media is more likely to be impacted by characteristics such as user-generated content, trends, and features of the platforms (Sashi et al. 2019). Comprehending how consumers engage with social media platforms and anticipating their choices will be important for brands to adjust their communication and marketing techniques accordingly.

2.7 Significance of Social Media in the Environment of Modern Business

In the modern business environment, social media has become an integral part of each brand, deeply influencing the international and competitive landscape of the corporation (Islam et al. 2019). This review of the literature digs into the evolving significance of social media in the business aspect of today's world, the evolution of Internet access, and how brand owners are implanting it within social media to maximise economic advantages.

The evolving significance of social media in today’s competitive and globalised corporate landscape: the evolving importance of social media in today’s modern landscape of business could be imputed to its capability of connecting businesses with a diverse and broad audience. As per Busalim and Ghabban (2021), businesses are required to reach further with contemporary marketing techniques as the world becomes increasingly globalized. Platforms on social media give businesses a mechanism for engagement with consumers globally, breaking down geographical boundaries and encouraging multinational opportunities for businesses. Therefore, social media has changed the consumer interaction procedure with businesses, highlighting authenticity, direct interaction, and transparency. Companies that successfully leverage the platforms of social media could develop a robust relationship with their customers, promote loyalty to the brand, and achieve a competitive advantage within the market.

Development of Internet access and its possible effect on social media use The development of Internet access has played an important role in powering the widespread adaptation of platforms for social media (Mason et al., 2021). As the usage of the Internet constantly evolves, more individuals have access to the platforms of social media. This direction has significant importance for brands, as they could now reach a wider potential customer base. The usage of social media platforms has become an essential part of individuals’ daily lives as the availability of mobile devices and reasonable data plans increase. Platforms of social media act as a basic source of entertainment, information, and interaction for individuals from various geographical and demographic locations.

Figure 4: Benefits of Social Media Marketing

(Source: Cheung et al. 2019, P-125)

Business owners are funding social media to increase economic advantages. Owners of businesses are increasingly acknowledging social media marketing's economic potential. As the online presence of consumers evolves, companies comprehend the requirement to develop a robust online presence to stay competitive and relevant. With an investment in social media marketing, companies could:

1. **Target particular demographics:** Platforms on social media give progressive options for targeting, allowing companies to guide their marketing efforts at particular demographic locations and interests. This targeted method enhances the effectiveness of marketing campaigns and increases investment returns.
2. **Improve brand visibility:** social media enables companies to reach a broader audience and enhance the visibility of the brand (Cheung et al. 2019). Businesses could position themselves as leaders of the industry and draw possible consumers by making compelling and engaging content for their followers.

3. **Drive sales and generate leads:** platforms on social media suggest brands have different possibilities for generating leads and driving sales. From shippable elements to sponsored posts, companies could guide consumers smoothly from purchase to discovery.

4. **Enhance customer engagement:** social media encourages real-time communication with consumers, enabling brands to answer queries, give personalised assistance to customers, and address concerns. Positive interaction with customers promotes advocacy and loyalty to the brand.

2.8 Social Media Marketing as an Operational Tool

Marketing through social media has evolved from an unconventional method to become an indispensable and functional tool for companies. This review of the literature investigates the modification of social media marketing, emphasising its voyage from an uncommon technique to a functioning tool. In addition, it examines the important role of social media within a competitive atmosphere and entrepreneurship.

Modification of social media marketing from an uncommon to a functional method: in the early days of social media, the marketing of social media was seen as an experimental and unexplored technique, frequently viewed as an unconventional advertising method (Coates et al. 2019). Thus, as platforms on social media gained popularity and user engagement increased, companies identified the possibility of connecting with a broad and engaged audience base.

The research proposed that the majority of brands now use social media as an important tool for marketing, demonstrating the transition from experimentation to mainstream adaptation. Marketing through social media has established its efficiency in guiding customer engagement, awareness of the brand, and growth of revenue, generating its shift to an integral and functional element of modern strategies for marketing.

Figure 5: Social Media Marketing for Business

(Source: Jin et al. 2019, P-252)

The role of social media in a competitive environment and entrepreneurship

1. **Improving customer engagement:** social media allows real-time and direct interaction among consumers and businesses. Entrepreneurs could employ their consumers, collect feedback, and reply to queries promptly, promoting the trust and loyalty of the brand. Within a competitive environment, robust engagement with consumers could be a differentiating element for victory.
2. **Developing brand identity:** social media enables entrepreneurs to convey and form the identity of their brand efficiently. Via visually attractive content, genuine interactions, and consistent messaging, entrepreneurs could develop an uncommon persona for their brand, differentiating it from the competition.
3. **Encouraging small businesses and start-ups:** platforms of social media suggest an accessible and cost-efficient way for small businesses and start-ups to develop a digital presence and connect with their target audience (Jin et al. 2019). Businesses could reach a broader audience and compete with competitors that are

more established by leveraging social media advertising, compelling content, and resource-deprived ventures.

4. **Influencer marketing:** within competitive industries, influencers on social media could play an important role in promoting the credibility of the brand and reaching audiences. Cooperating with influencers could help entrepreneurs achieve visibility and expand their consumer base.

5. **Monitoring market trends and competitors:** platforms of social media give an important understanding of the activities of the competitors and the evolving trends of the market. Entrepreneurs could assess the social media strategies of their competitors and respond to the consumers and their content to inform their efforts for marketing and remain ahead of them.

2.9 Different Sentiment Analysis Level (SAL)

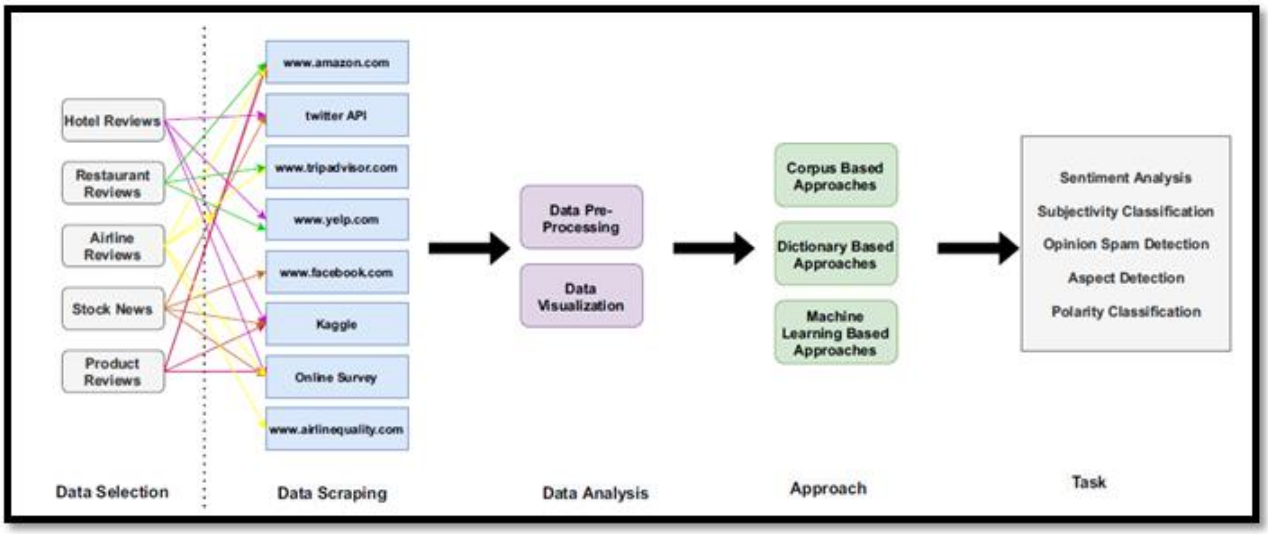


Figure 6: Basic steps on SAL procedure

(Source: Wankhade, Rao and Kulkarni, 2022)

Sentiment analysis on the basis of Document level –

A sort of sentiment analysis called document-level sentiment analysis looks at the entirety of a document and gives it a single polarity. Though it isn't frequently used, it

may be used to rank chapters or pages in a book as good, terrible, or neutral. For this level of classification, we can utilize supervised or unsupervised learning approaches. The two fundamental issues with document-level sentiment analysis are cross-domain and cross-language issues. It has been demonstrated that domain-specific analysis is extremely accurate while remaining highly domain-specific. A vocabulary set must be employed that is exclusive to one domain in order to complete these activities (Wankhade, Rao and Kulkarni, 2022).

Sentiment analysis on the basis of Sentence level –

At this stage of analysis, each sentence is examined to determine its polarity. This is especially helpful when a document is connected with a wide variety and mix of feelings. Subjective classification is the categorization level connected to this degree of analysis. With additional training data and processing capabilities, each sentence will be decided separately using the same techniques as the document level. The polarity of each phrase can be combined to determine the document's overall attitude or utilized individually. For certain applications, document-level sentiment analysis is insufficient. Finding subjective sentences has been the main emphasis of earlier research on sentence-level analysis, but increasingly challenging tasks, including those incorporating conditional phrases and ambiguous claims, need for a more thorough sentiment analysis (Asif et al. 2020).

Sentiment analysis on the basis of Phrase level –

Sentiment analysis is the technique of identifying and categorizing opinion words at the phrase level. There may be several elements to one sentence or simply one. A product evaluation of many lines, for instance, could include just one element. Recent years have seen a rise in interest in this subject among scholars. Sentence level analysis is preferable since a text may contain both positive and negative assertions, in contrast to document level analysis, which concentrates on categorizing the whole content as subjective, either positive or negative. The word is the most basic building block of language, and its polarity is directly related to how subjective the phrase or document is where it occurs. For instance, it is quite likely that an adjective-containing phrase is a

subjective statement. Additionally, the terminology used to communicate anything reflects the person's demographics, such as sex and age, as well as their desires, social standing, personalities, and other psychological and sociological traits. As a result, text sentiment analysis is based on words.

Sentiment analysis on the basis of Aspect level –

Sentiment analysis is done at the aspect level, and because each sentence may have numerous aspects, it is also known as aspect level sentiment analysis. All of the elements that are employed in the sentence are the major emphasis, and once each element has been given a polarity, the overall sentiment of the statement has been computed (Piedrahita-Valdes et al. 2021).

2.10 Review on Data Collection

Data collecting from the internet is necessary for the first step of sentiment analysis, and it may be done using a variety of methods, including site scraping and social networking. Furthermore, data may be gathered from news channels, e-commerce websites, forums, blogs, and other websites. Text data may be combined with other data sources, such as video and audio recordings, location data, and more, depending on the task's objectives, using sentiment analysis (Nemes and Kiss, 2021).

1. **Some kinds of Social media websites like facebook, twitter and instragam:** Social media platforms are used to gather social data. It demonstrates how users engage with a product through information access, sharing, and exchange. Social media is frequently used in research on people, groups, and behavior as a source of dynamic data. Internet programs that are web- or mobile-based and let users produce, view, and share user-generated material are referred to as social data.
2. **Subjective Forums:** Users may converse on message boards about a range of subjects, post their ideas and opinions, and send text messages to ask for assistance. As a result of the fluidity of user-generated information, forums are an intriguing source for sentiment analysis. Furthermore, by using forums as a resource, researchers may carry out sentiment analysis on a certain subject.

3. **Some blogging websites like Trustpilot and Review.io:** An overview of a viewpoint, some data, or some links can be found in a few paragraphs in a short blog post. These are referred to as "websites" collectively, and they are organized chronologically, with the most recent item appearing first, much like similar kind research paper. A great tool for doing sentiment analysis on a variety of topics is blogs.

4. **Other Commercial Websites like Amazon:** A website without a dedicated review site is a prime instance of an e-commerce site where customers may leave ratings and comments on a company or organization. A professional assessment site may have a professional review area, for instance, while an e-commerce site might have a product review section. In this instance, a descriptive research was carried out to classify the various airline services.

2.11 Chapter Summary

This review of the literature sheds light on the evolving value of social media within the modern business environment. It investigates how social media has developed from an uncommon method of marketing to becoming an indispensable and functional tool for brands. This review highlights the modification of interaction between consumers and businesses due to the widespread adaptation of social media platforms, which has unlocked new challenges and opportunities for businesses. The main findings showcase that social media gives companies enhanced exposure to an international audience, allowing for improved brand recognition and sales. Actual-time interaction with consumers promotes loyalty and trust, whereas access to their information gives an important understanding of their preferences and behaviors. Marketing through social media demonstrates itself to be a targeted and economical way to reach particular demographics and efficiently foster businesses. Harnessing and understanding the possible effect of social media on brands is important for staying relevant and competitive in today's interconnected and globalised world. Brands needed to creatively and strategically utilise social media to stand out among the growing competition and engage with consumers in a meaningful way.

As the study progresses, the review of the literature emphasises the requirement for brands to adjust to the changing behaviour and perceptions of consumers on social media platforms. Accommodating and anticipating the future trends of customers would be important for efficient marketing strategies on social media. Transitioning to the methodology of the study and output, the study will examine the usage of social media for promotion and marketing, its effect on key implementation indicators, and practical strategies for enhancing outcomes. It would also forecast future customer needs, assess sentiment analysis and models of predictive analysis, and assess the importance of social media for consumers and businesses.

Chapter 3: Methodology

Web Scraping Tools

Web scratching is the practice of collecting information from websites using a computerised handle. A few websites expressly disallow clients from utilising computerised instruments such as those used in this instructional exercise to scrape their information. Websites may do this for two reasons (Mitchell, 2018):

1. They have a legitimate reason for ensuring their information. For example, Google Maps does not permit clients to ask as many questions as they come about in a brief period of time.
2. Repeated demands on a website's server may devour transfer speed, slow down other clients, and possibly overburden the server to the point where it ceases to reply at all.

It is critical to check the worthy utilisation arrangement of the target site to guarantee that getting to their site with computerised apparatus does not abuse their terms of use. In legal terms, web scratching against a website's wishes may be a dark range.

Web scraping tools can be found in the Python standard library in the form of the urllib package. Urllib provides tools for handling URLs, particularly the urlopen() function, which can be used to open URLs within a programme (Lawson, 2015).

```
>>> from urllib.request import urlopen
webpage = urlopen(url)
```

In order to access the HTML content of the page, the HTML page must first be read using the read() function of the HTML page object. The read() function returns a series of bytes. Subsequently, the HTML page object must be decoded using the function decode(). The bytes must be converted into a string using the UTF-8 encoding.

```
>>> read_byte = webpage.read()
>>> htmlpage = read_byte.decode("utf-8")
```

Utilizing string methods, it is possible to extract data from the HTML of a web page. For example, the.find() function can be used to search the text of HTML tags for the desired title and extract it. To begin, the desired title can be extracted from the web page specified in the above example. If the index for the initial character in the title and the index for the final character in the closing tag are known, then the title can be extracted using a string slice. Additionally, the index for the opening tag 'title' can be obtained by passing a string 'title' to the '.find() function.

```
soup = BeautifulSoup(htmlpage)
gettext = soup.get_text()
```

Beautiful Soup is a program that Python uses to get information from HTML and XML documents. It helps us easily navigate, search, and change the parse tree using our preferred parser. This often helps developers save lots of time and effort.

Regular expressions are patterns that can be used to find text in a string. Regular expressions, or regex for short, are supported by Python through the re module of the standard library (Hajba, 2018).

```
>>> import re
```

Metacharacters are special characters used in regular expressions to indicate various patterns. For example, an asterisk (*) indicates zero or more occurrences of whatever is immediately preceding the asterisk.

This technique allows for scraping data from websites that regularly update their information. However, it is important to note that multiple requests for a page in a short period of time can be interpreted as suspicious or malicious use of a website.

Sentiment Analysis by using of python library

Big data fundamentally entails sentiment analysis. It entails the capacity to ascertain the feelings, views, or attitudes of an entity—a thing, a person, a group, or a service—through a range of written materials. Incorporating machine learning into documents to extract new variables and relationships in data that enable more complex analysis of collected information is unavoidably required in order to include sentiment analysis in papers. Predictions that were previously unable to be proven or were unreliable can now be provided to a business user as a consequence of the new connections created. This document's main objective is to use a variety of analytic approaches to determine if the current data stream contains positive or negative sentiment (Ouyang et al. 2015, October).

There are many different stakeholders involved in the subject of study known as sentiment analysis, including businesses, organisations, and governments. It is a useful tool for learning about people's perspectives on a range of issues, including goods, rules, and laws. Sentiment research has the potential to provide businesses with a previously unthinkable competitive edge when done properly. Sentiment analysis may be used to spot and prepare for certain social events, such as strikes, secessions, and uprisings, in a manner similar to social media monitoring. Any of the following groups might encompass the uses and issues that sentiment analysis aims to solve (Rani and Kumar, 2019).

1. **Opinion Assessment of Items and Administrations:** It's perhaps the most direct and basic sentiment analysis available. It's a fantastic method for businesses to see how customers feel about their goods without having to do a tonne of surveys, like customer satisfaction surveys. In this manner, businesses may determine if customers enjoy a specific product based on what they write on discussion forums, blogs, or social media platforms. This gives businesses a competitive edge since they can always see if customers enjoy their products or not, and if not, they may adjust their strategy as quickly as feasible (Rani and Kumar, 2019).

2. **Correction Process in Opinions:** Many individuals share their opinions on things on online purchasing platforms. The majority of the time, we can post a review, but occasionally we can also rate anything. We may rate a product from one to five stars, for instance. However, occasionally you don't assign the proper rating. Our remarks can be examined by a sentiment analysis system, which can identify the issue and offer a solution (Rani and Kumar, 2019).

3. **Item suggestion in systems advancement:** An online business may prioritise the goods that it offers depending on customer ratings; for instance, it may not promote items with unfavourable feedback.

4. **What's the best way to position online ads?** Advertisers for certain items may occasionally demand that their ads only appear on internet pages with positive content in order to avoid appearing on sites with unfavourable material.

5. **Political involvement and standing:** Sentiment analysis has the ability to show how people feel about a variety of political issues. These subjects may cover, but are not limited to: a certain political party; a certain candidate; public perceptions of political parties; certain government policies implemented; and modifications to these policies through time.

6. **Financial market analysis:** On the basis of the information available on websites, forums, and social media platforms about a particular company, the market development of that company can be inferred from the value added by the juxtaposition of all the opposing views (Rani and Kumar, 2019).

NLTK : TextBlob

The text processing library, TextBlob, is a powerful tool for performing Natural Language Processing (NLP) tasks in Python. It is based on two well-known Python libraries, NLTK (Natural Language Testing) and Pattern (Pattern). TextBlob enables users to combine the functionality of both of these libraries in a user-friendly interface. TextBlob can be used for a variety of NLP tasks, including morphological analysis and entity extraction. Additionally, it can be used for sentiment analysis and machine translation (Wang et al. 2016, August).

Algorithms for sentiment analysis typically focus on the identification of beliefs, sentiments, and emoticons within a set of texts. However, the extent of established sentiments can vary significantly depending on the method used. A standard analyzer typically only allows for the identification of three fundamental polar emotions, namely positive, negative, and neutral. However, more sophisticated models are able to go beyond this and identify six more universal emotions (such as anger, annoyance, apprehension, fear, joy, sorrow, and surprise).

In addition, depending on the subject matter of the work, it is possible to extract additional information from the context, such as the author or a subject matter that, upon further investigation, can avoid a more intricate problem than a standard polarity classification, namely, the identification of subjectivity or objectivity (Wang et al. 2016, August).

The TextBlob package adopts a rule-based approach to sentiment analysis, which necessitates the provision of a pre-defined collection of categorised words. Such words can be uploaded, for instance, from the National Language Translation Knowledge (NLTK) database. Furthermore, sentiments are defined on the basis of semantic relationships and the occurrence of each word in the input sentence, which enables a more accurate output. The user can receive the sentiment scores in terms of polarity and subjective after completing the first step and feeding the necessary input data into the Python model.

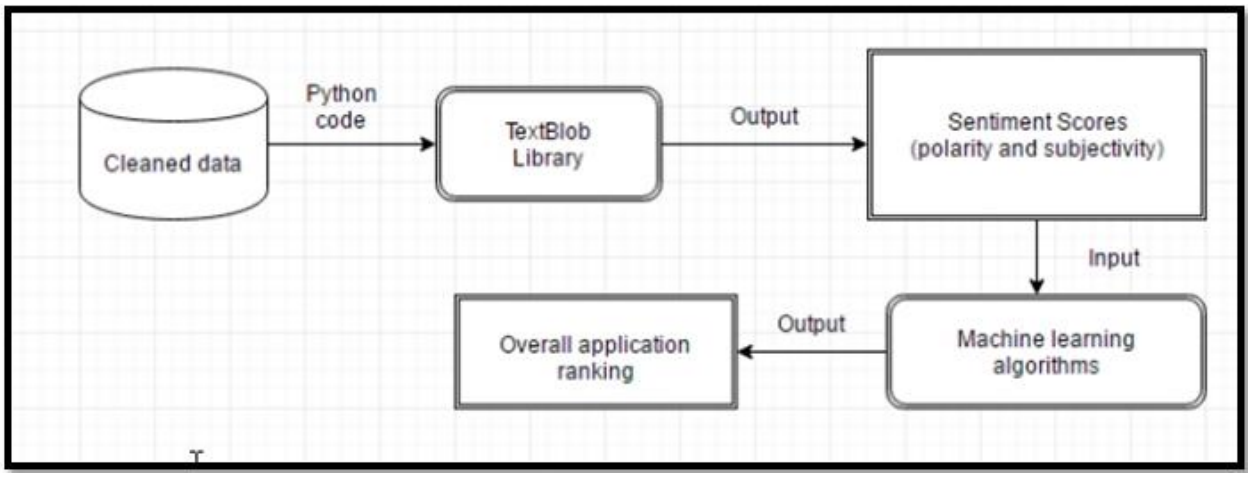


Figure 7: Sentiment Analysis by using TextBlob and machine learning

(Source: Created by Author)

The output of TextBlob for a polarity task is a float with a range of [-1.0 to 1.0], where -1.0 is negative and 1.0 is positive. As the statement doesn't contain any terms from the training set, the score can alternatively be set to 0, which denotes a neutral evaluation of the statement. In contrast, the outcome of a subjective/objective identification job is a float in the range [0.0 to 1.0], with 0.0 denoting a statement that is extremely subjective and 1.0 denoting one that is highly objective. The TextBlob sentiment analyzer may be used in conjunction with a wide range of Python programs, including models based on Kaggle datasets like movie reviews and Twitter API-based sentiment analysis of tweets (Ouyang et al. 2015, October).

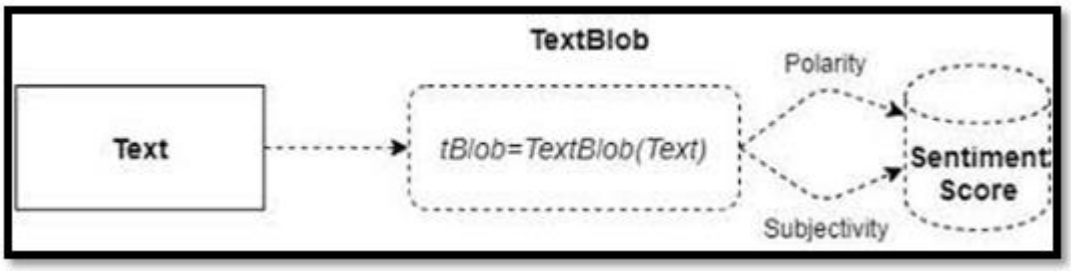


Figure 8: Two features of TextBlob for sentiment scores

(Source: Created by Author)

Polarity and subjectivity, the two primary metrics of polarity and subjectivity, are crucial for sentiment analysis. The term "polarity" describes the depth or intensity of an emotion communicated in text that reflects the person's behavior. A person's perspective or point of view on a certain subject or idea is referred to as their subjectivity. A subjective statement might simply be a claim presented in the form of an opinion; it does not always denote polarity or portray a behavior. A phrase is entered into the text-based library function TextBlob(), which returns the subjectivity and polarity of the text. Plots were created for each date using the mean of the two data points. The way this function works is shown in the above figure. One of the numerous components utilized in the machine learning and deep learning domains to address a variety of issues is NLP (natural language processing). NLP includes sentiment analysis as a discipline. Sentiment analysis is the technique of applying natural language processing (NLP) to analyse natural language in order to understand the public's sentiment in an area of interest (Rani and Kumar, 2019).

Textblob may be used for a number of tasks, including the extraction of word phrases, the application of speech tagging and tokenization, the extraction of word inflections, the application of lemmatization, and WordNet integration. It is mostly used for sentiment analysis and has an internal classifier that has already been trained. It can also do multiple sentiment assessments.

- **Rule-based sentiment analysis**

Rule-based sentiment analysis is one of the most fundamental techniques for determining text sentiment. The preparation needed for this procedure is minimal, and the idea behind it is simple. The sentiment of a text is determined using this method without the use of any machine learning techniques. By counting the number of times a user used the word "sad" in a tweet, for example, we may determine the emotion of a statement (Wang et al. 2016, August).

```
from textblob import TextBlob

WordText = TextBlob("The item was great!")
print(WordText.sentiment)
```

```
Sentiment(polarity=1.0, subjectivity=0.75)
```

Unrecognized words will be disregarded by Textblob. The algorithm will take into account words and phrases to which it can attribute polarity, as well as average values, to determine the final score.

Another widely used rule-based sentiment analysis tool is the Valence-Aware Dictionary for Sentiment Reasoning (VADER). It utilises a set of lexical features, such as words that are classified as positive or negative based on their semantic orientation, to calculate the sentiment of a text. VADER returns the likelihood that a given input sentence is likely to be positive, negative, or neutral (Wang et al. 2016, August).

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
textanalyzer = SentimentIntensityAnalyzer()
words = "The food was very nice!"
check = textanalyzer.polarity_scores(words)
print("{:-<60} {}".format(words, str(check)))
```

The food was very nice! ---------- {'neg': 0.0, 'neu': 0.542, 'pos': 0.458, 'compound': 0.5244}

Vader is made to work well with social media data and, when paired with data from Twitter, Facebook, etc., may produce fruitful outcomes. However, the main drawback of this rule-based sentiment analysis approach is that it ignores context and just concentrates on individual words. Instances like "the party was wild" won't be considered by any token-based algorithms.

- **Embedding based models**

In the realm of Natural Language Processing (NLP), a form of word representation known as embedding is employed. In embeddings, synonyms are depicted by similar vectors that, when positioned in an N-dimensional space, tend to be closely situated. Within the context of NLP, this type of text representation finds application in Python tools that utilize embeddings to infer the sentiment of text. This approach not only enhances the representation of Natural Language Programming (NLP) text but also contributes to the augmentation of model performance. Flair stands as a notable example of such a tool.

Flair serves as an accessible framework for modern Natural Language Processing (NLP) tasks. It offers versatile capabilities, encompassing functions such as embedding texts, pre-sentiment analysis, and Named Entity Recognition (NER) extractions. This tool, as evidenced by Ouyang et al. (2015, October), plays a pivotal role in bolstering the efficiency and effectiveness of NLP text representation and analysis.

```
from flair.models import TextClassifier
```

```
from flair.data import Sentence

model = TextClassifier.load('en-sentiment')
words = Sentence('The place is nice!')
model.predict(words)
print('For the sentence : ', words.labels)
```

For the sentence : ['Sentence[5]: "The place is nice!"'/'POSITIVE' (0.9831)]

- **Pattern-based analysis**

A text processing component of Pattern is included, and it may be used to mine text data (such as web crawls, etc.). A text can be sent to the sentiment() method, which produces a tuple with the text's subjectivity and polarity. Between -1 and +1 are the polarities, while 0 and 1 are subjectivity (Ouyang et al. 2015, October).

- **How to calculate the ratio between subjectivity and polarity**

A wide range of human activities have been made easier by the use of information technology. One such activity is the growth of IT in the public reviews, where information and access are accessible to submit reviews of activities of company on websites and social media platforms. This study uses web text mining tools to look at how polarity, subjectivity, and clustering relate to online visitor evaluations. The acquired data from internet reviews are analyzed by the research using sentiment analysis. The group based analysis is also done by ratio analysis between subjectivity and polarity values. Researcher may have design by calculating through any formula to find out the best solution. Hence ratio between polarity and subjectivity stands as shown in below equation.

$$Ratio\ for\ positive\ score = \frac{Positive\ words}{Subjectivity\ of\ the\ respective\ sentence}$$

$$Ratio\ for\ negative\ score = \frac{Negative\ words}{Subjectivity\ of\ the\ respective\ sentence}$$

Positive score and Negative Score can be obtained be the above formula and these score indicates about the magnitude sign of the Polarity. In another way sentiment value can be designed through the multiplication of both factor polarity and subjectivity which can be reliable to find out the best solution by labelling the value in manual definition.

- **How to find the best satisfaction level from the selected companies review**

Company's Name	Positive Words	Negative Words	Total Webscraped Sentences	Positive Remark per sentences	Negative Remark per sentences	Customer's Best Satisfaction

From the above table it can be cleared that the necessary columns are indicating about selecting the best company according to positive remark per sentence and negative remark per sentences which can be calculated as shown in below:

$$\textbf{Positive Remark per sentences} = \frac{\textbf{Total Positive words}}{\textbf{Total Webscraped sentences}}$$

$$\textbf{Negative Remark per sentences} = \frac{\textbf{Total Negative words}}{\textbf{Total Webscraped sentences}}$$

The best company is that who have highest Positive Remark per sentences and lowest Negative Remark per sentences. Hence comparative analysis in markets can be done in such type of methodology which can help to determine which company is best according to their homogeneous groups.

Chapter 4: Data Analysis and Discussion

Prior to delving into the technical analysis, it is crucial to provide an overview of the proposed methodology and how sentiment analysis will be developed for small businesses such as Bettys, Yorkshire Soap Company, Black Yak, Masons Yorkshire Gin, Harrogate Candle Company, and others. For the fulfilment of the project's objectives, it is necessary to retrieve the comments by people from the reviewed websites and blogging websites like Trustpilot and review.co.uk, etc., and again, the collection of retrieved comments by web scraping can lead to the further analysis of sentiments, which tells about people's emotions and satisfactions with this company. And also, the further improvement of the company's needs is dependent on this sentiment analysis. So the whole technical analysis is very interesting in how we deal with people's comments about taking action on the improvement of small business-oriented companies. Now we will also shortly see our proposed steps for technical analysis.

Step 1: Retrive the comments from the review websites or blogging websites by using web scraping tools in Python.

Step 2: After web scraping, perform sentence cleaning, tokenization, and visualising the word cloud by word frequencies.

Step 3: Sentiment analysis by combined and average methods through the pattern and Textblob with NLTK (Natural Language Toolkit, which is NLP, i.e., Natural Language Process-oriented).

Step 4: Creating a dataset table saved in the format of CSV for classifying the sentences into two parts, i.e., subjectivity and polarity It is briefly said that one's opinion focuses on his or her comment's remarkable sentiment or emotion. Also show the scatter characteristics between two variables, i.e., subjectivity and polarity, which are two major characteristics of the TextBlob and Pattern Python libraries.

Step 5: Separately identify the positive and negative words that can impact small businesses.

Step 6: Visualise the word cloud based on positive and negative words.

Step 7: Find the sentences that contain those positive and negative words according to the sorting for identifying the key terms for the requirements of the company's improvements.

Step 8: Also apply steps 1 to 7 for the other four companies, and the whole result is combined to be stored as a grand csv file for classification.

Step 9: However, the grand dataset comes from the primary analysis taken by textblob and pattern, and after this dataset is post-classified and fed onto the CNN layers,

Step 10: Final conclusion with remarks is to be made on the basis of Sentiment score analysis.

Let's look at the steps one by one according to technical improvements.

```
url_1="https://www.trustpilot.com/review/www.bettys.co.uk/"
url_2="https://www.trustpilot.com/review/yorkshiresoap.co.uk/"
url_3="https://www.reviews.co.uk/company-reviews/store/black-yak/"
url_4="https://uk.trustpilot.com/review/masonsofyorkshire.com/"
url_5="https://uk.trustpilot.com/review/harrogateorganics.co.uk?page=77/"
```

Initially, we set up the URL addresses of review websites according to Bettys, Yorkshire Soap Company, Black Yak, Masons Yorkshire Gin, and Harrogate Candle companies.

Then, we imported some web scraping and sentiment-oriented libraries.

```
pip install pattern
from pattern.en import sentiment
import requests
import nltk
```

```
from bs4 import BeautifulSoup
from wordcloud import WordCloud
import spacy
nlp = spacy.load('en_core_web_sm')
from textblob import TextBlob
```

The Pattern Library is used for implementing natural language processing (NLP). From the pattern library, the sentiment function key is imported for the language English (especially parts of speech in English). The request modules are used in Python for sending the HTTP requests. This is handy for doing things like using web APIs or scraping data from websites. The most suitable NLP library for sentiment analysis will vary based on our project needs. However, it is suggested to use three highly recommended sentiment analysis libraries: TextBlob, NLTK, and VADER. Beautiful Soup is a programme that Python uses to get information from HTML and XML documents. It helps us easily navigate, search, and change the parse tree using our preferred parser. This often helps developers save lots of time and effort. A wordcloud is a way to show how often words appear in a text. The bigger the word, the more times it appears. Spacy can help create systems that extract information, understand human language, or prepare text for deep learning.

Web Scraping AI Tool

```
req=requests.get(url_1)
req.encoding = 'utf-8'
htmlpg = req.text
print(htmlpg)
soup = BeautifulSoup(htmlpg)
gettext = soup.get_text()
```

Cleaning the Text in such a way removing \n, \ and / with a white space

```
txt_clean= gettext.replace("\n", " ")
```

```
txt_clean= txt_clean.replace("/", " ")
txt_clean= ''.join([txt for txt in txt_clean if txt != "\'"])
txt_clean
```

NLP PHASE

```
line_sentence=[]
wordtokens = nlp(txt_clean)
for i in wordtokens.sents:
    line_sentence.append((i.text.strip()))
```

Sentiment analysis is a type of text classification that falls under natural language processing (NLP). In simpler terms, sentiment analysis is when a piece of writing is sorted into different feelings like positive or negative, happy, sad, or neutral. Subjectivity classification is when we determine whether a sentence is expressing an opinion or not. The opinionated sentences that are classified as expressing positive or negative opinions are known as sentence-level sentiment classification. The main thing about sentiment analysis is to analyse a bunch of words to understand the feelings expressed by them. Usually, we measure this feeling with a score of good or bad, which is called polarity. We can understand if the overall feeling expressed is good, okay, or bad by looking at the polarity score. The Natural Language Toolkit (NLTK) comes with a list of common words called stop words. These words, like "a", "an", "the", "of", and "in", are often used but don't carry much meaning on their own. The stopwords in nltk are the words that appear frequently in data. These are words that we should avoid using when talking about our content. These things are already set and cannot be taken away.

Sentiment Analysis using Textblob

```
senti_textblob=[]


for sent in line_sentence:
    txt_sent= TextBlob(sent)
    sent_polarity= txt_sent.sentiment.polarity
    sent_subjectivity= txt_sent.sentiment.subjectivity
```

```
senti_textblob.append([sent,sent_polarity,sent_subjectivity])
```

Polarity have 3 magnitude i.e. -ve, zero and +ve but Subjectivity stands for opinion or not opinion (start from zero to +ve)

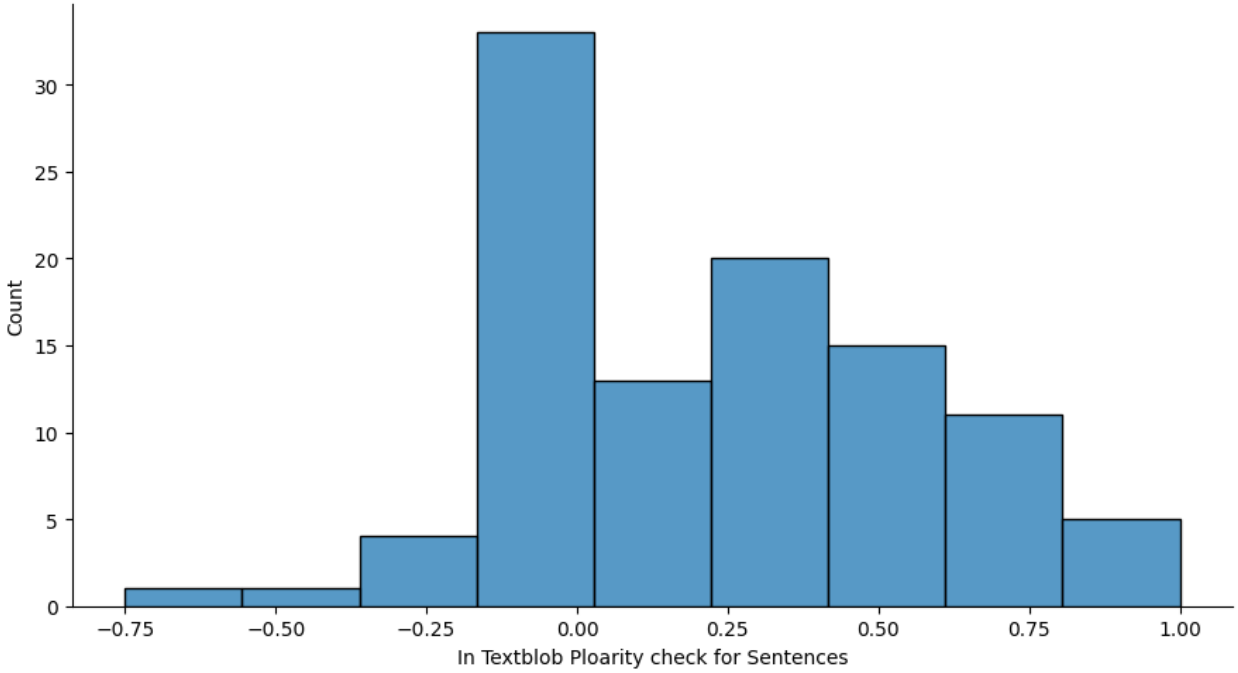


Figure 9: In TextBlob Poalrity graph visualize the sentences that how sentences are look like in terms of people's feelings

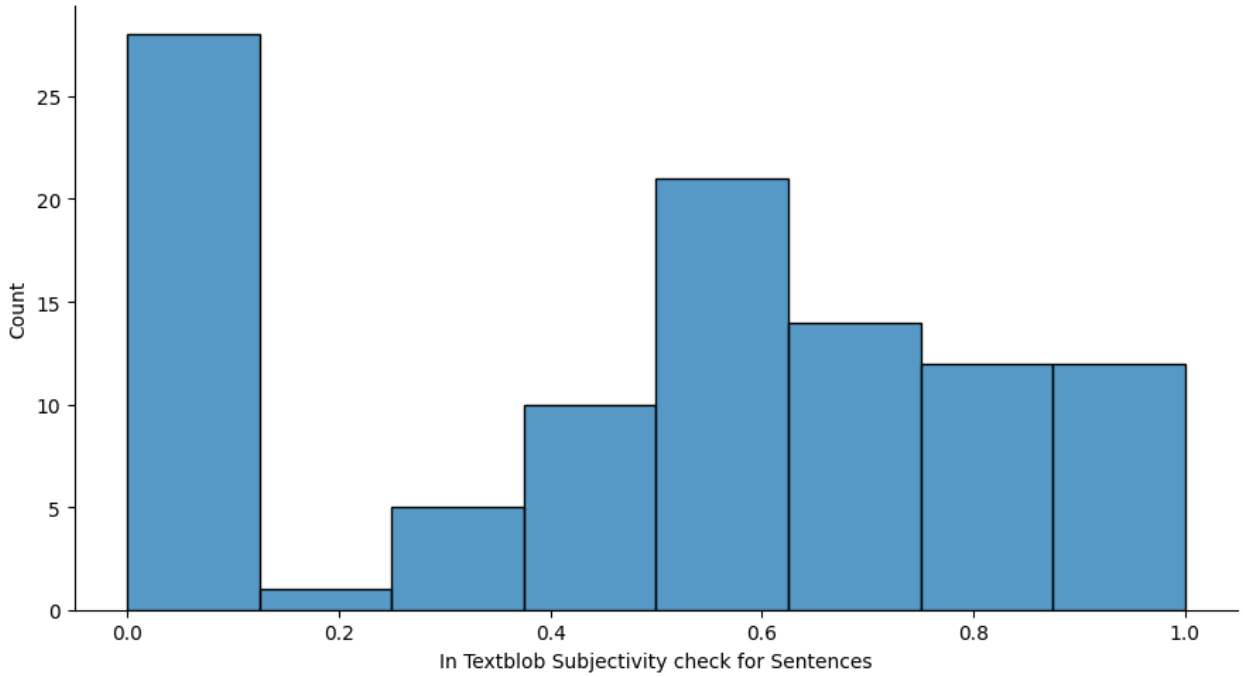


Figure 10: In TextBlob subjectivity graph visualize the sentences that how sentences are look like in terms of people's opinions

Similarly we also did the sentence visualizations by the Pattern analysis. The graphs from this analysis are quite similar as compared with textblob analysis.

	Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2
0	Bettys Reviews | Read Customer Service Reviews...	0.000000	0.000000	0.000000	0.000000
1	Excellent4.7VERIFIED COMPANYIn the Cake Shop c...	0.159253	0.507684	0.159253	0.507684
2	The procedure for placing the order and paymen...	0.487500	0.487500	0.487500	0.487500
3	The large cake was in a strong box but a large...	0.290476	0.397619	0.290476	0.397619
4	The staff managed to find some plastic sacks t...	0.100000	0.400000	0.100000	0.400000
...	...	...	...	...	...
98	We verify companies and reviewersVerification ...	0.200000	0.300000	0.200000	0.300000
99	We advocate against biasIt goes against our gu...	0.000000	0.000000	0.000000	0.000000
100	We also ensure all reviews are published witho...	0.000000	0.000000	0.000000	0.000000
101	Take a closer lookBettysare you human?Choose c...	0.000000	0.000000	0.000000	0.000000
102	All rights reserved.	0.000000	0.000000	0.000000	0.000000

103 rows × 5 columns

From the above figure, we can see that the table contains some columns like index, sentence, polarity1, subjectivity1, polarity2, subjectivity2. Polarity 1, Subjectivity 1, comes

from sentiment analysis using textblob, and Polarity 2, Subjectivity 2, comes from sentiment analysis using pattern.

How we categorise each sentence by the sentiment label is important, and it is shown in the next steps.

Average Polarity and Subjectivity Calculation

```
csv_df['Polarity_mean']=round((csv_df['Polarity1']+csv_df['Polarity2'])/2,2)
csv_df['Subjectivity_mean']=round((csv_df['Subjectivity1']+csv_df['Subjectivity2'])/2,2)
csv_df
```

Sentiment values are achieved from the multiplication factors between two separate analysis to identify the sentiment values.

```
csv_df['Sentiment_value']=round(csv_df['Polarity_mean']*csv_df['Subjectivity_mean'],2)
```

Table will look after this implementation which is shown below.

Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_value
Bettys Reviews | Read Customer Service Reviews...	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
Excellent4.7VERIFIED COMPANYIn the Cake Shop c...	0.159253	0.507684	0.159253	0.507684	0.16	0.51	0.08
The procedure for placing the order and paymen...	0.487500	0.487500	0.487500	0.487500	0.49	0.49	0.24
The large cake was in a strong box but a large...	0.290476	0.397619	0.290476	0.397619	0.29	0.40	0.12
The staff managed to find some plastic sacks t...	0.100000	0.400000	0.100000	0.400000	0.10	0.40	0.04

Hence sentiment Values are categorized in the range(-0.5, 0.5). And we have shown the categorized mapping values in the format of non-numerical names.

-3: very worst

-2 : worst

-1 : dislike

0 : No sentiment

1 : like

2 : better

3 : great

Now look the table how it is:

```
conditions = [
    (csv_df['Sentiment_value'] < -0.5),
    (csv_df['Sentiment_value'] >= -0.3) & (csv_df['Sentiment_value'] < -0.1),
    (csv_df['Sentiment_value'] >= -0.1) & (csv_df['Sentiment_value'] < 0),
    (csv_df['Sentiment_value'] ==0),
    (csv_df['Sentiment_value'] > 0) & (csv_df['Sentiment_value'] < 0.3),
    (csv_df['Sentiment_value'] >= 0.3) & (csv_df['Sentiment_value'] < 0.5),
    (csv_df['Sentiment_value'] >= 0.5)
]

values = ['-3','-2','-1','0','1','2','3']

csv_df['Sentiment'] = nmp.select(conditions, values)

csv_df
```

Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_value	Sentiment
ews | Read ner Service Reviews...	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00	0
'VERIFIED PANYIn the e Shop c...	0.159253	0.507684	0.159253	0.507684	0.16	0.51	0.08	1
ocedure for e order and paymen...	0.487500	0.487500	0.487500	0.487500	0.49	0.49	0.24	1
ake was in g box but a large...	0.290476	0.397619	0.290476	0.397619	0.29	0.40	0.12	1
nanaged to ome plastic sacks t...	0.100000	0.400000	0.100000	0.400000	0.10	0.40	0.04	1
...	...	...	...	...	...	...	...	...
companies and Verification	0.200000	0.300000	0.200000	0.300000	0.20	0.30	0.06	1

103 rows × 9 columns

- Select those sentences whose sentiment values are negative

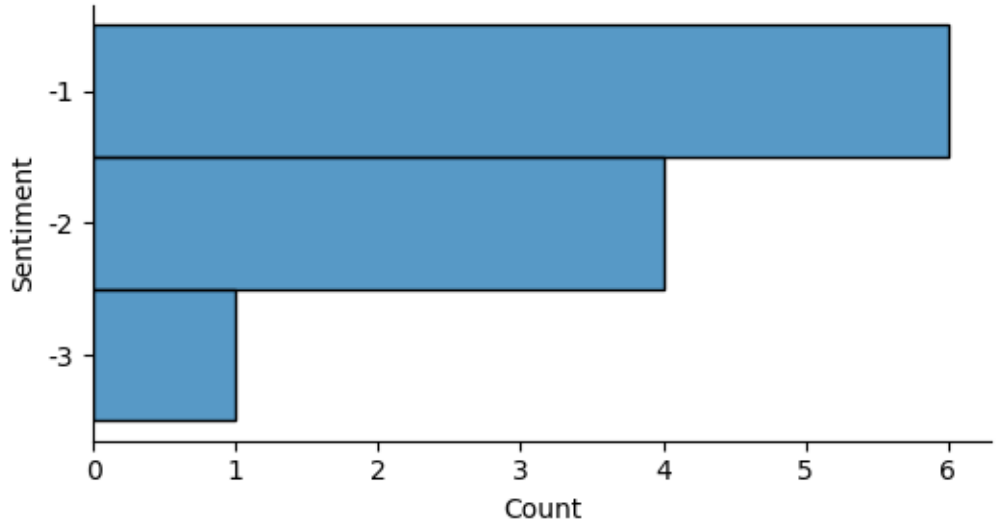


Figure 11: Ratio among the negative commented sentences within its range lies between -1 to -3.

- Select those sentences whose sentiment values are positive

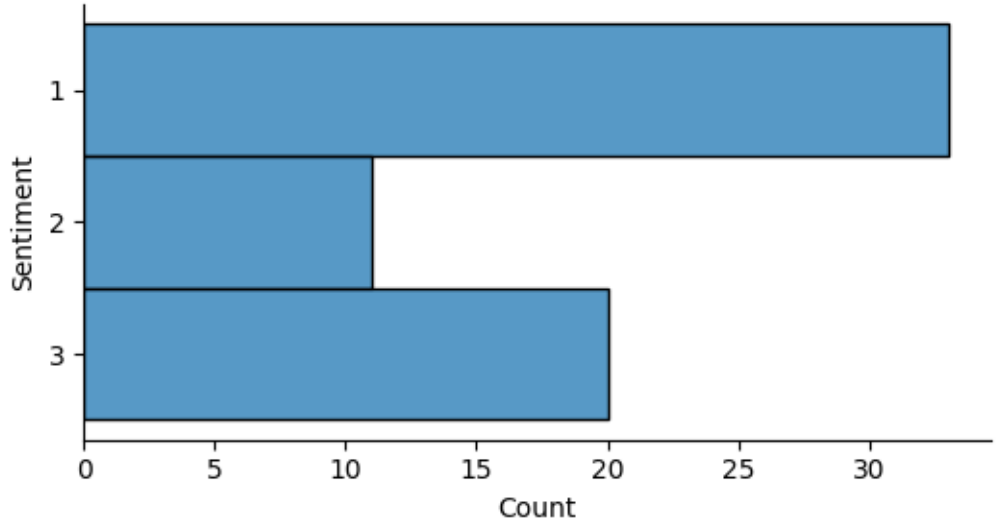


Figure 12: Ratio among the positive commented sentences within its range lies between +1 to +3.

- Visualization: Distribution plot in Word Frequency and Word Cloud

Tokenizer in Regular Expression

```
tokenizer = nltk.tokenize.RegexpTokenizer('\w+')
tokens = tokenizer.tokenize(txt_clean)
```

```
tokens[0:50]
```

```
['Bettys',
 'Reviews',
 'Read',
 'Customer',
 'Service',
 'Reviews',
 'of',
 'www',
 'bettys',
 'co',
 'ukCategoriesBlogLog',
 'inFor',
 'businessesFor',
 'businessesLog',
 'inCategoriesBlogFood',
 'Beverages',
 'TobaccoBakery',
 'PastryCake',
 'ShopBettysOverviewReviewsAboutBettys',
 'Reviews',
 '20',
 '068',
 'Excellent4',
 '7VERIFIED',
 'COMPANYIn',
 'the',
 'Cake',
```

Token words are revised by stopwords

```
import nltk
nltk.download("stopwords")
token_modified = []
for sep_word in tokens:
   if sep_word not in stopwords:
      token_modified.append(sep_word)
```

```
token_modified[0:50]
```

```
['Bettys',
 'Reviews',
 'Read',
 'Customer',
 'Service',
 'Reviews',
 'www',
 'bettys',
 'co',
 'ukCategoriesBlogLog',
 'inFor',
 'businessesFor',
 'businessesLog',
 'inCategoriesBlogFood',
 'Beverages',
 'TobaccoBakery',
 'PastryCake',
 'ShopBettysOverviewReviewsAboutBettys',
 'Reviews',
 '20',
 '068',
 'Excellent4',
 '7VERIFIED',
 'COMPANYIn',
 'Cake',
 'Shop',
 'categorywww',
 'bettys',
 'co',
 'ukVisit',
```

- Check the frequency of most occurrence of words:

```
token_count=pnd.value_counts(nmp.array(token_modified))
```

```
token_count.head(10)
```

```
experience    27
days          21
August        17
Date          17
We            16
Bettys        15
reviews       13
time          11
fancies       11
much          10
dtype: int64
```

Distribution Plot

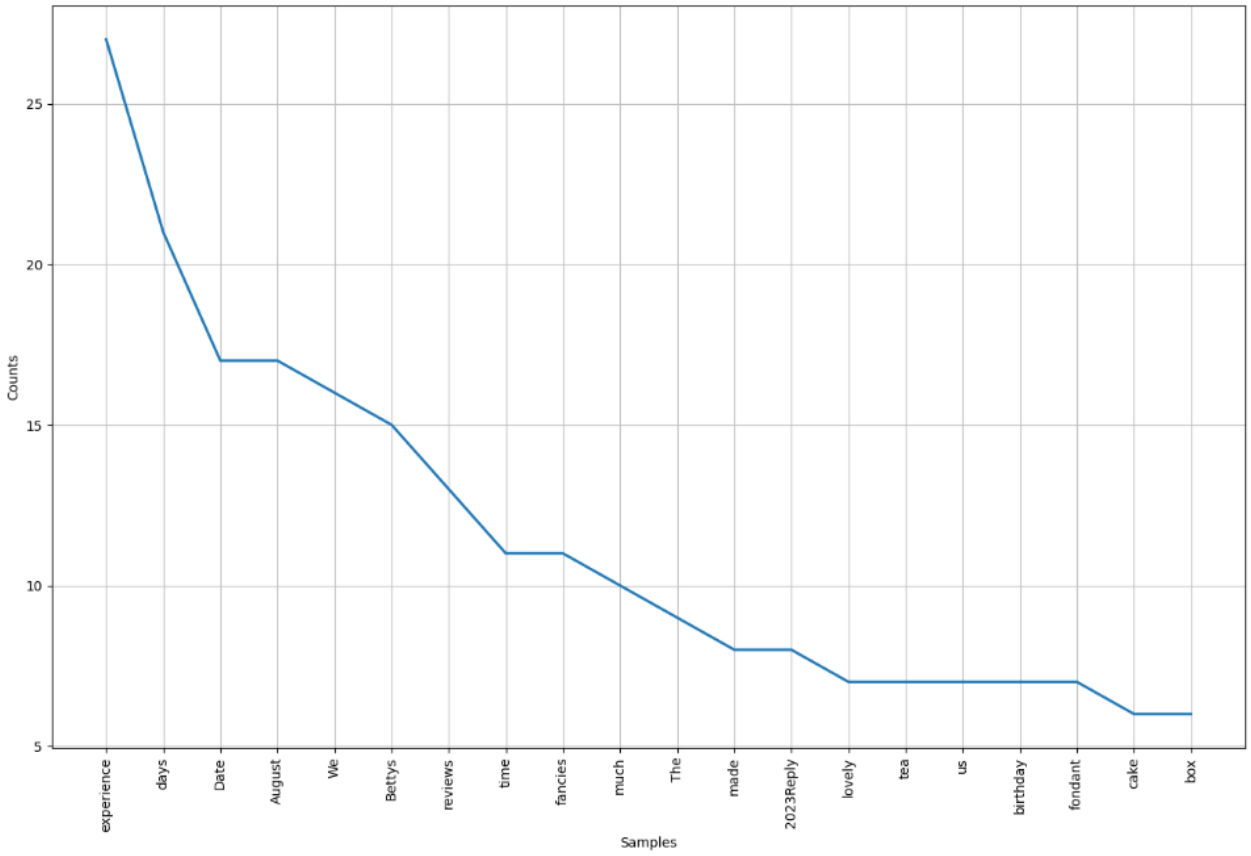


Figure 13: Distribution Plot for all tokenization by stopwords rectification

From the Figure 13, it reveals that the distribution plot of most occurrence words is important to show and illustrate how the words in both sentiments are repeatedly used.

- Visualization for positive words and negative words

Highest, Middle and Least words to show the potentiality of positive words

```
[ ]  res_highest_pos=pos_token_count.index[0]
     res_mid_pos=pos_token_count.index[int(len(pos_token_count)/2)]
     res_lowest_pos=pos_token_count.index[int(len(pos_token_count)-1)]
```

```
[ ]  res_highest_pos

     'experience'
```

```
[ ]  res_mid_pos

     'Bakery'
```

```
[ ]  res_lowest_pos

     'write'
```

Select those sentences by most and least occuring postive word

```
[ ]  pos_rev_highest = select_data_for_pos[select_data_for_pos['Sentence'].str.contains(res_highest_pos)]
     pos_rev_highest
```

	Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_
12	agoAfternoon teaFour of us went for afternoon ...	0.316667	0.683333	0.316667	0.683333	0.32	0.68	
14	Four very happy customersDate of experience: ...	0.500000	0.583333	0.500000	0.583333	0.50	0.58	
18	What a wonderful experience we had too, thorou...	0.500000	0.816667	0.500000	0.816667	0.50	0.82	
20	As for the staff they couldnt do enough to mak...	0.305729	0.652083	0.305729	0.652083	0.31	0.65	
21	Cannot recommend Bettys tea room enough, thank...	0.325000	0.750000	0.325000	0.750000	0.32	0.75	
23	Date of experience: August 11, 2023JFJoy Falcu...	0.400000	0.900000	0.400000	0.900000	0.40	0.90	
	Date of experience:							

```
[ ]  pos_rev_least = select_data_for_pos[select_data_for_pos['Sentence'].str.contains(res_lowest_pos)]
     pos_rev_least
```

	Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment
98	We verify companies and reviewersVerification ...	0.2	0.3	0.2	0.3	0.2	0.3	

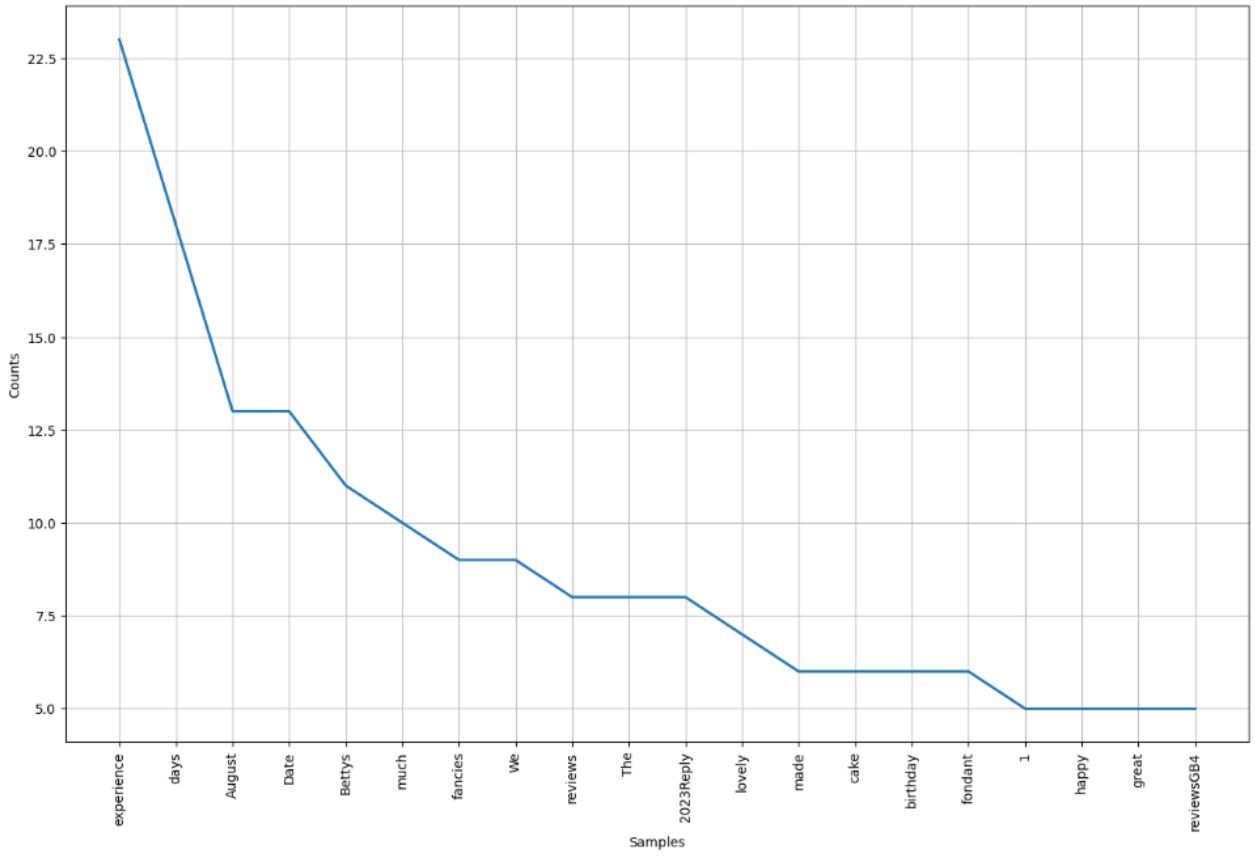


Figure 14: Distribution Plot for all positive words

From the Figure 14, it reveals that most and least occurring positive sentiment oriented words are shown at their counts. From the distribution plot it also helps for searching those words which suggest about some specific characteristics.

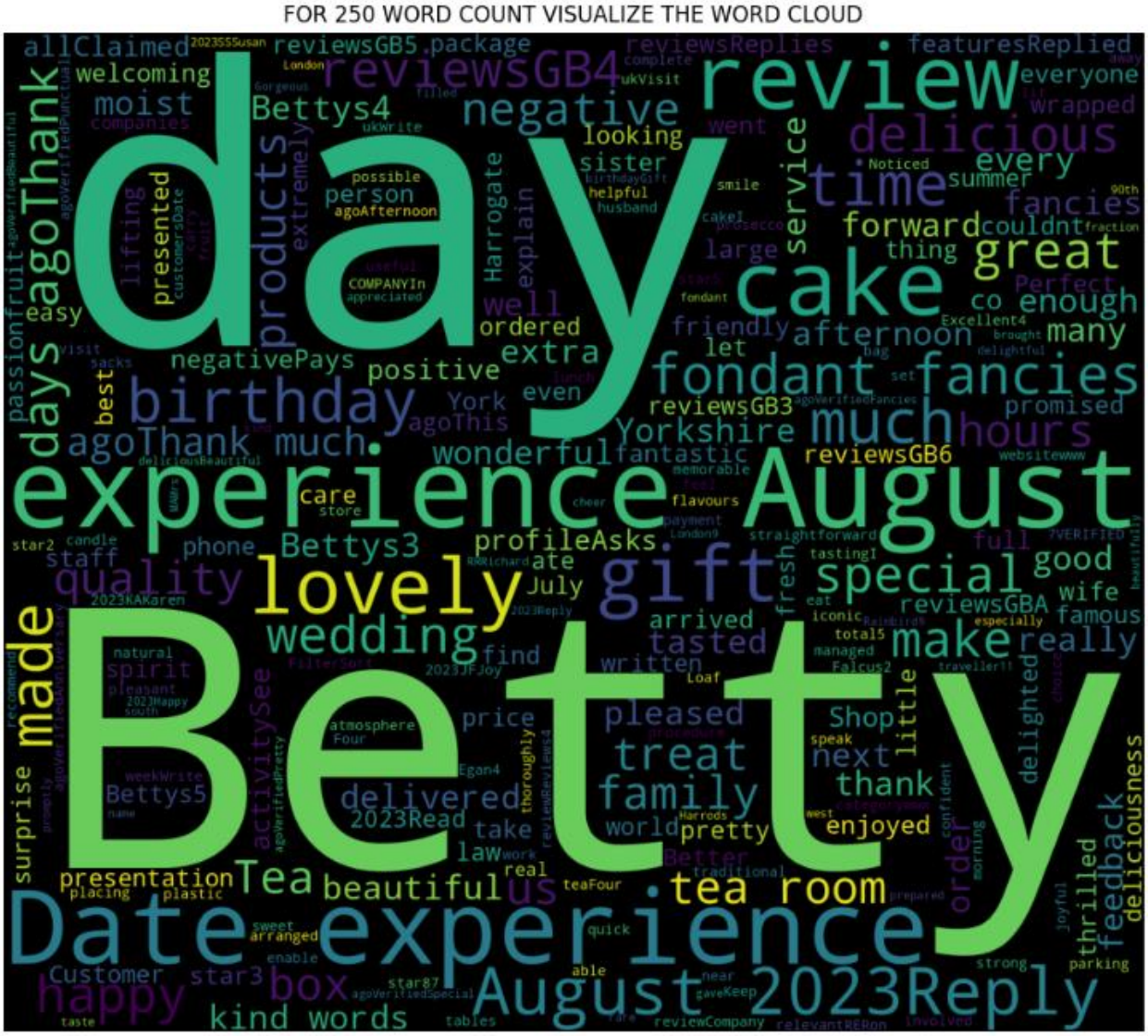


Figure 15: WordCloud for positive words

From the Figure 15, it reveals that some high font sized words (here experience, lovely, day, Betty,....) mostly occurred, and low font-sized words exist through least occurred in the sentences.

- For negative Reviews

For negative reviews, it also suggest as similar as positive words findings.

```
tokenizer = nltk.tokenize.RegexpTokenizer('\w+')
negtokens = tokenizer.tokenize(negwords)
stopwords = nltk.corpus.stopwords.words('english')

neg_token_modified = []
for sep_word in negtokens:
    if sep_word not in stopwords:
        neg_token_modified.append(sep_word)

neg_token_count=pnd.value_counts(nmp.array(neg_token_modified))
neg_token_count
```

```
time          5
perfection    2
chocolate     2
We            2
present       2
             ..
mail          1
Royal         1
weeks         1
2             1
platform      1
Length: 93, dtype: int64
```

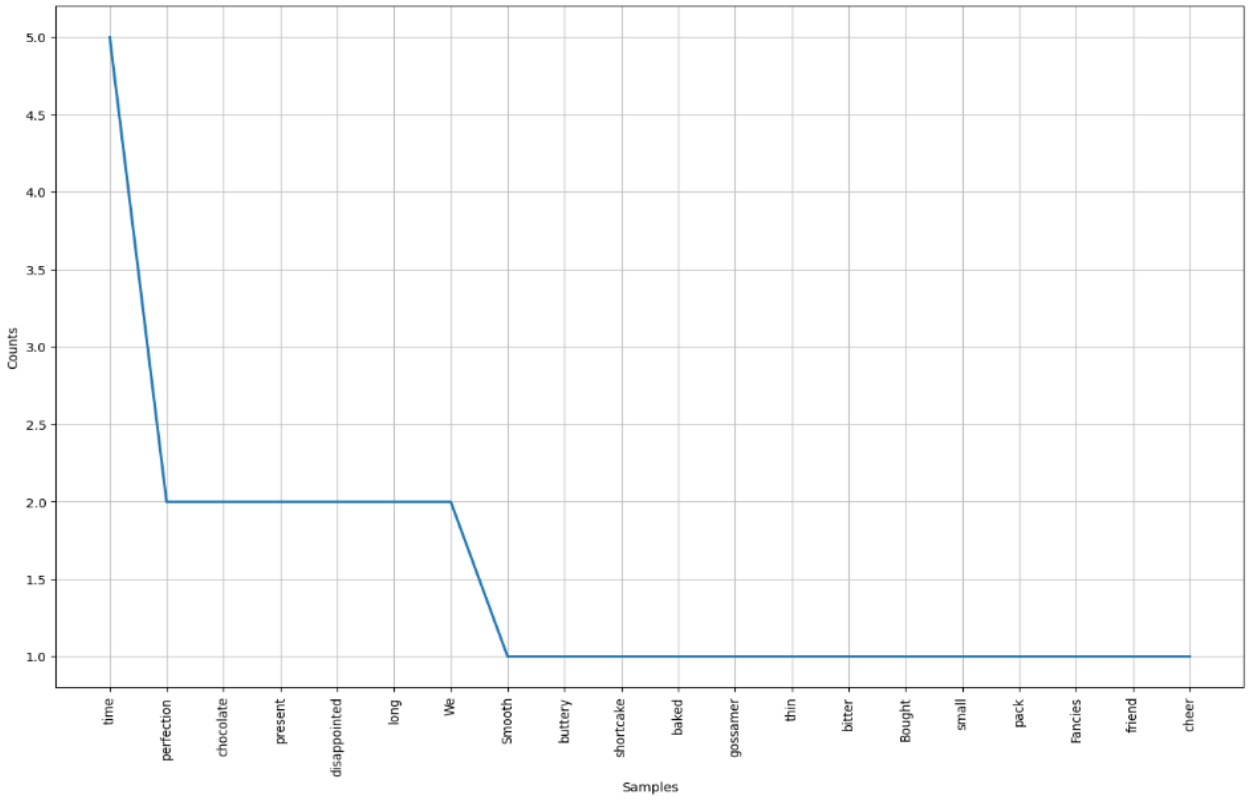


Figure 16: Distribution Plot for all negative words

As similar Figure 14, the above figure also justifies for negative words.

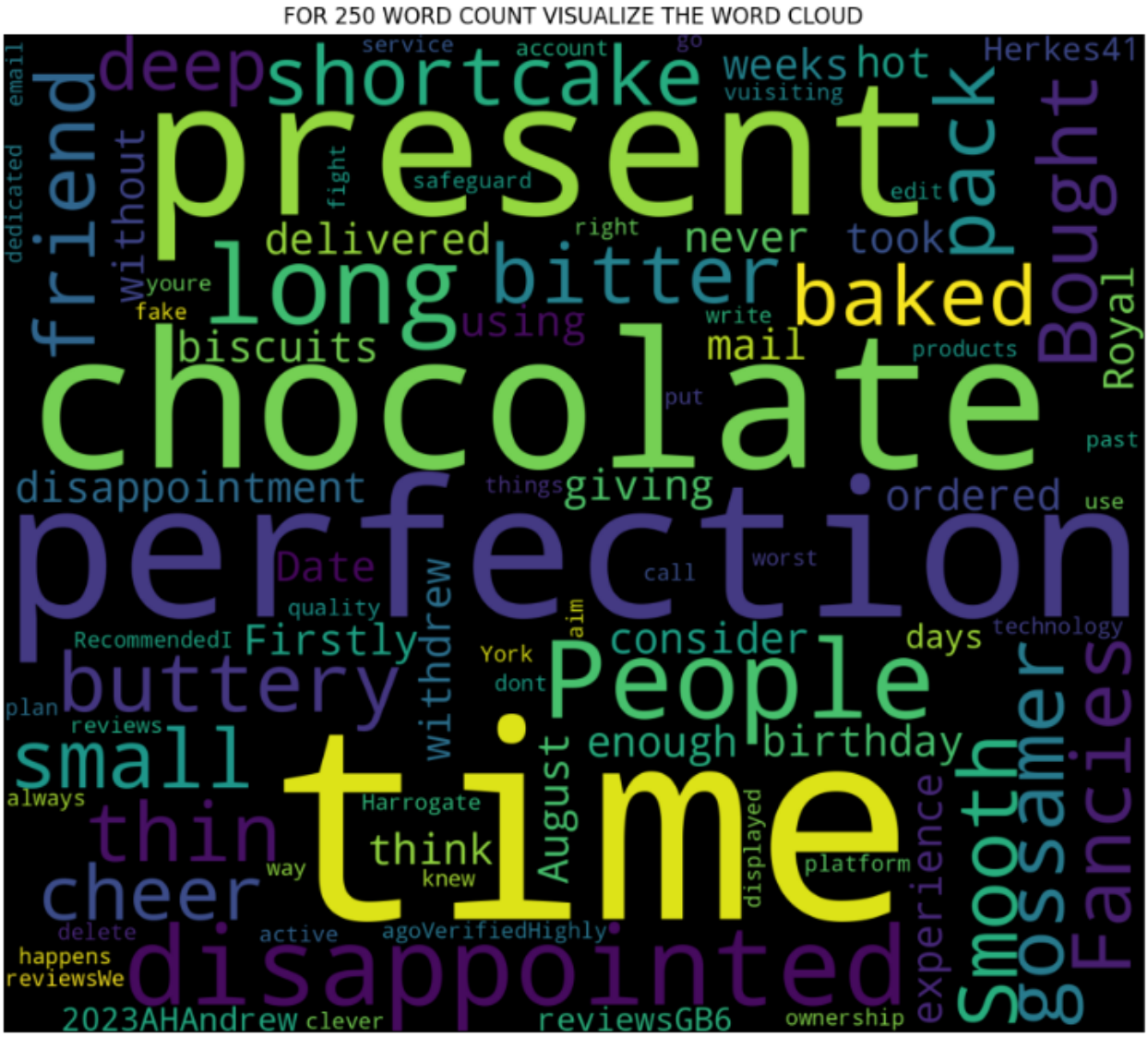


Figure 17: WordCloud for all negative words

From the Figure 17, it reveals that some high font sized words i.e. chocolate, long, bitter, baked, ...etc. are existed for negative sentiments here.

Highest, Middle and Least words are shown for potentiality to negative words

```
[ ]  res_highest_neg=neg_token_count.index[0]
     res_mid_neg=neg_token_count.index[int(len(neg_token_count)/2)]
     res_lowest_neg=neg_token_count.index[int(len(neg_token_count)-1)]
```

```
[ ]  res_highest_neg
```

```
'time'
```

```
[ ]  res_mid_neg
```

```
'email'
```

```
[ ]  res_lowest_neg
```

```
'platform'
```

Select those sentences which contain most and least occurring negative words

```
[ ]  neg_rev_top = select_data_for_neg[select_data_for_neg['Sentence'].str.contains(res_highest_neg)]
```

```
[ ]  neg_rev_top
```

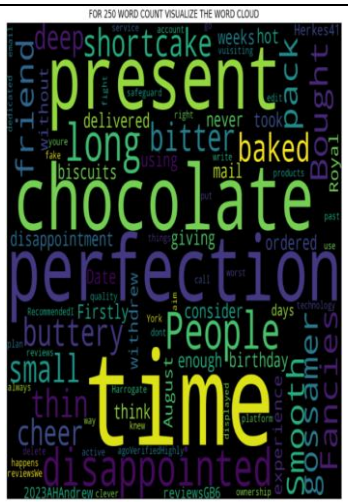
	Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_value	Senti
42	Firstly you withdrew the hot chocolate i had o...	-0.042857	0.404762	-0.042857	0.404762	-0.04	0.40	-0.02	
83	We aim for perfection but from time to time th...	-0.357143	0.767857	-0.357143	0.767857	-0.36	0.77	-0.28	
90	People who write reviews have ownership to edi...	-0.091667	0.500000	-0.091667	0.500000	-0.09	0.50	-0.04	

```
neg_rev_least = select_data_for_neg[select_data_for_neg['Sentence'].str.contains(res_lowest_neg)]
neg_rev_least
```

	Sentence	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_value	Senti
94	We fight fake reviewsWe use dedicated people a...	-0.166667	0.916667	-0.166667	0.916667	-0.17	0.92	-0.16	

Conclusion on this "Bettys" Company: We have tested the reviews collected from the review websites and we depict that some mostly occurred positive and negative words can demand the company's requirement for further improvement by which the customer's satisfaction can be optimized and regulated.

Similarly we have observed for other four companies. We have summarized the word cloud for all companies as respectively in positive and negative sense.

Company ’s Name	Positive Word Cloud	Negative Word Cloud	Comments
Bettys			As seen in two types of word cloud, it can be depict that 1. Problem in quality of item: like chocolate,butter and cheer’s amount in the cake 2. serving the item: Slow food serving is another reason to occur problem. These are two necessary to improve for this company.
Yorkshire Soap Company			As seen in two types of word cloud, it can be depict that 1. Problem in quality of Soap product: like chocolate,butter and cheer’s amount in the cake

			This is essential for this firm to flourish.
Black Yak			As seen in two types of word cloud, it can be depict that 1. Slow serving the item: To advance for this business, this is required.
Masons Yorkshire Gin			As seen in two types of word cloud, it can be depicted that 1. Problem in quality of item: lime is a keyword in the word cloud referes to a problem occuring in the product of the company For this firm to develop, this is important.

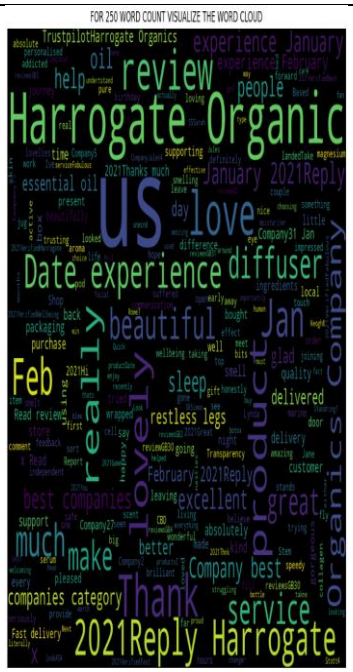
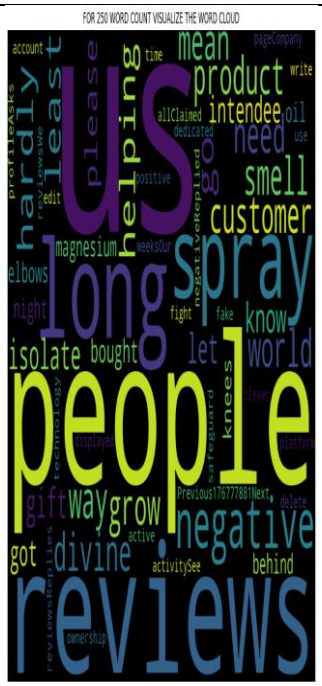
Harrogate Candle			As seen in two types of word cloud, it can be depict that 1. Candle Quality in duration Question about hard candle or soft candle is arised from seen in negative words from the word cloud. This is required for this firm to advance.

Table 1: Review analysis through positive and negative sentiments oriented Word cloud and it's regarding comments on improvements of the respective companies

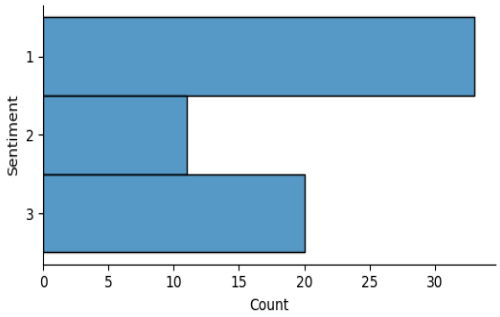
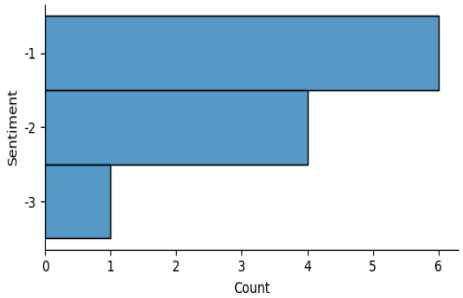
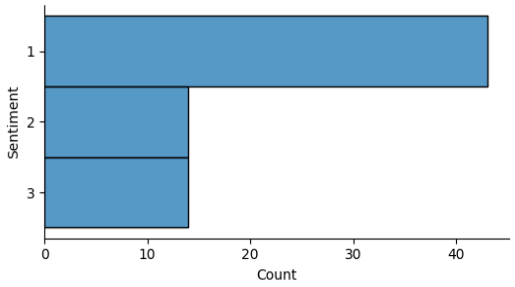
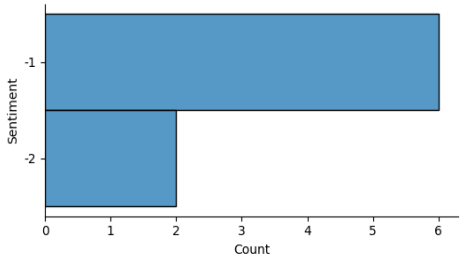
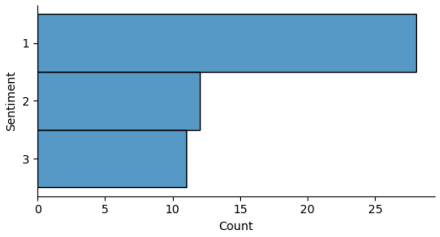
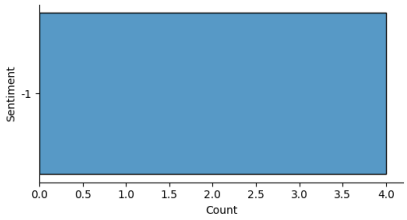
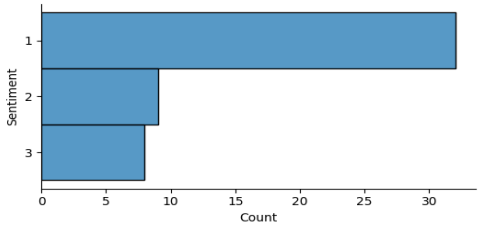
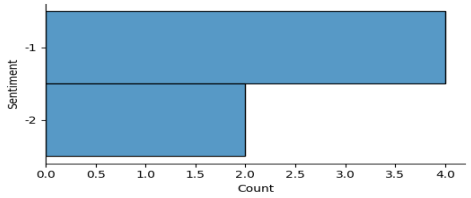
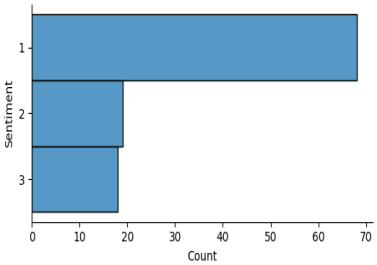
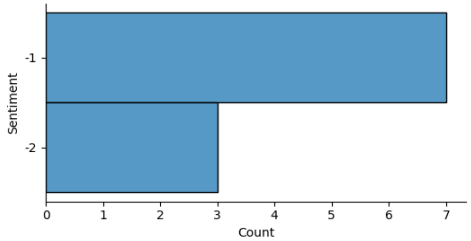
Company	Positive sentiment plot	Negative sentiment plot
Bettys		
Yorkshire Soap Company		
Black Yak		
Masons Yorkshire Gin		
Harrogate Candle		

Table 2: Visualization of Positive and Negative Sentiments throughout the sentiment analysis

Company's Name	Positive Words	Negative Words	Total Webscraped Sentences	Positive Remark per sentences	Negative Remark per sentences
Bettys	507	93	103	4.92	0.18
Yorkshire Soap Company	515	105	120	4.29	0.2
Black Yak	326	88	71	4.59	0.27
Masons Yorkshire Gin	518	57	84	6.17	0.11
Harrogate Candle	609	69	176	3.46	0.11

Table 3: Review analysis through positive and negative words and its remark per sentence

From the above table, we depict that Masons Yorkshire Gin is the best satisfied company by the customer's review.

All company's positive and negative sentiment words are collected in separately and saved in the format of CSV file. Likewise positive_words_company1.csv and negative_words_company1.csv file refers to the company of "Bettys".

- Merge all csv files into one csv file and show some of new Grand CSV dataset's Characteristics

[554 rows x 10 columns]

	Unnamed: 0	Polarity1	Subjectivity1	Polarity2	Subjectivity2	Polarity_mean	Subjectivity_mean	Sentiment_value
count	554.000000	554.000000	554.000000	554.000000	554.000000	554.000000	554.000000	554.000000
mean	60.945848	0.252746	0.427156	0.252829	0.428238	0.252708	0.427780	0.171047
std	42.153778	0.311838	0.328174	0.312480	0.328774	0.312077	0.328472	0.254979
min	0.000000	-0.750000	0.000000	-0.750000	0.000000	-0.750000	0.000000	-0.560000
25%	27.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	55.000000	0.200000	0.500000	0.200000	0.500000	0.200000	0.500000	0.090000
75%	86.000000	0.472917	0.666667	0.478333	0.675244	0.480000	0.670000	0.280000
max	175.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Total Sentiment Counts

```
[ ]  new_dataset.Sentiment.value_counts()
```

```
 1    208
 0    181
 2     69
 3     61
-1     23
-2     11
-3      1
Name: Sentiment, dtype: int64
```

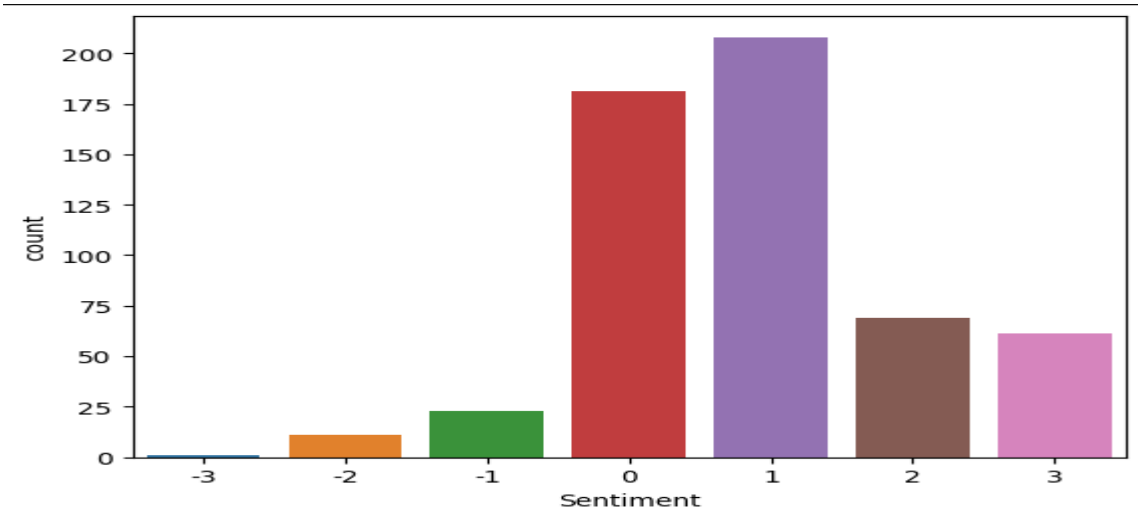


Figure 18: Label of total Sentiment class from all company's review

Discussion: We have seen from the whole technical analysis that web scraping from review or blogging websites is so much more efficient for retrieving web contents like people's comments, and therefore collecting the sentences is ready for an analysis of sentiments. In this technique, we have analysed multiclass sentiments, whose range lies between -3 and +3, i.e., seven categories of sentiments. The average calculation of polarity and sentiments is estimated for providing the sentiment values, and the table is stored in a CSV format on the drive. All technical analysis was done on the Python-based Google Colab platform. We have recognised the word cloud and word frequencies for the selected five companies, and hence, we estimate the customer's best satisfaction among the selected five companies. Our experiment is best suited for only sentiment analysis by both combined and average methods from predefined Python libraries, i.e., textBlob and pattern. Requirements of fulfilling demands by the customer's satisfaction can be achieved by textBlob and pattern analysis for this dataset.

Survey taken through Polling:

The survey has given useful information on what customers think about their interactions with companies on social media sites. It is clear that a large number of customers often connect with brands on social media and enjoy the conversations and material offered. However, both good and bad experiences have been recorded, showing that there is space for improvement in how corporations use social media.

Customers' levels of confidence in the information and marketing materials offered by companies vary, suggesting that organisations should prioritise being transparent and trustworthy. Additionally, differing customers' perceptions of personalised content and offers underscore companies' need to find a balance that speaks to their demographic.

Chapter 5: Conclusion and Future Recommendations

This study wanted to determine if it was possible to use a combination of different algorithms to get a consensus on what people said about trending tweets in Twitter API data streams. We used textblob, the Lexicon approach, and labelled a dataset to test the results. The combination of the two algorithms was the most accurate. We think our combination method is both fast enough for the data stream and accurate enough for analysis. We conclude that our combination analysis is appropriate and timely for our data stream and accurate for sentiment analysis. It has been observed that web scraping from review or blog websites is not suitable for retrieving web content such as people's comments, and thus, collecting the sentences is not suitable for sentiment analysis. This technical analysis has been conducted using multiclass sentiments, which range between -3 and +3, i.e., 7 sentiment categories. The average calculation of the polarity and sentiment values is estimated, and the table is stored in a CSV format on disk. All of the technical analysis was conducted in Python on the Google Colab platform. For the five companies selected, the word cloud and word frequencies were identified, and thus the customer's best satisfaction was estimated. It is concluded that this experiment is best suited for sentiment analysis only, using both the combined method and the average method from the predefined Python libraries (TextBlob and Pattern).

Recommendations for improving the Sentiment Analysis:

Sentiment analysis transforms unstructured information into actionable insights. It is driven by machine learning, which is a subset of AI that relies on algorithms that drive the software to learn and improve itself through experience continually. Unstructured data, such as listening to social media, chatting with chatbots, transcripts from medical and call centres, survey responses, and so on, can be a powerful source of intelligence. These insights can be used to guide targeted, strategic changes to improve a brand's value proposition, customer experience, value delivery, and more. The results of sentiment analysis can be used to bring precise change and transformation to a company's business.

These changes can be made to areas that are causing the most negative sentiment (e.g., product features, pricing, return policy, customer service, or prices) or areas that are creating the most positive sentiment (net banking, price matching, etc.). These strategic measures can help businesses become more competitive, attract new customers, retain existing customers, sell more products and services, reduce customer servicing, make customers more profitable, improve marketing messages and campaigns, and more. Examples of sentiment analysis in various industries with practical applications demonstrate that the benefits of sentiment monitoring are essential to any modern organization. To summarise, in addition to the areas previously discussed, the three main areas where sentiment monitoring can generate significant returns for businesses are:

1. The next frontier in business is the ability to forecast the future through the use of highly accurate artificial intelligence. From forecasting the decline of oil prices due to the emergence of political instability in a particular region to predicting which television programmes will be popular in the markets for an on-demand content platform such as Netflix or the BBC, the range of market trends that sentiment analysis can encompass is endless. Take a look at this example of TikTok trend analysis in the clothing retail sector using a video content analysis tool.
2. In order to gain an advantage over competitors, it is essential to have comprehensive information about how they are impacting the target company available at any point in time. Through the use of sentiment analysis in business, companies can identify gaps in their marketing strategy, manage their brand reputation, and identify key areas where customer sentiment is either positive or negative. By utilising sentiment analysis, companies can focus on improving audience engagement, contextualising and granulating key performance indicators (KPIs), and creating better messaging for marketing and advertising campaigns.

3. In order to build and sustain a positive brand perception, it is essential to understand and empathise with customers. Sentiment monitoring provides a platform for a company to do just this. It assists a business in its influencer

marketing campaigns by selecting the appropriate channels. A case in point is Unilever's Dove Real Beauty Campaign, which was based on insights from thousands of surveys. The aim of the campaign was to create a positive perception of the brand as one that was inclusive, positive, and supportive of women and children of all colours, sizes, and ages. Dove's marketing campaign sought to challenge the rigid and unrealistic beauty standards imposed by beauty brands or fashion magazines and instead focused on building the self-esteem of everyday people. This campaign was a global success, resulting in an increase in public opinion, brand loyalty, and customer base.

4. Ultimately, businesses can only expand when they have a deep understanding of the people utilizing their products and services. This, in itself, is a significant challenge, given the complexity of the human experience, which is characterized by a vast array of complex emotions and behaviors. Artificial Intelligence provides us with the ability to not only separate these emotions from one another, but also to establish a threshold value on which to apply this emotional intelligence. With Repustate, this intelligence can be brought to the surface through a straightforward, user-friendly sentiment analysis dashboard. Not only does the dashboard allow businesses to view data, but it also tracks and simplifies complex data sets. Additionally, the monitoring tool can provide businesses with a quick view of trends by transforming the data into graphs, charts, and tables.

Key Framework design for user friendly:

It is undeniable that sentiment analysis has a significant role to play in business. Sentiment analysis can be a game-changer for the complete revitalization of a brand. To run a successful business with sentiment data, it is essential to be able to utilize the unstructured information for meaningful insights. For the past few years, machine learning models have largely relied on the manually generated features prior to classification, which has served this purpose well. However, the advantages of deep learning are that it automatically

extracts the relevant features, helps to remove redundant features, and eliminates the need for manual crafting.

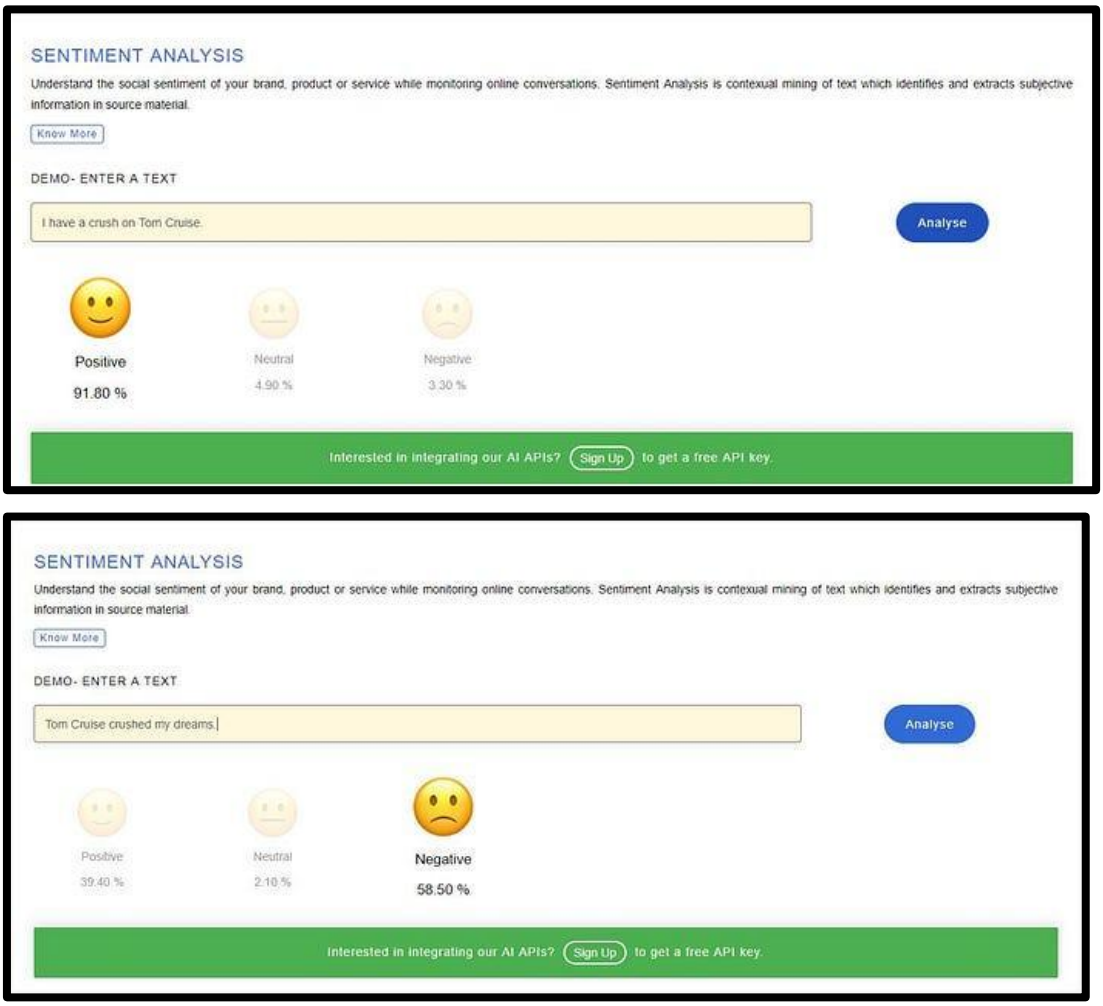


Figure 19: Framework tells about the sentiment label class

How does the improvised framework impact the company's improvement according to the sentiment analysis:

Sentiment analysis offers incremental benefits due to its ability to provide an in-depth understanding of evolving market trends and consumer preferences, regardless of the sector in which the company operates. Emotional mining of audience experience data derived from various sources, including social media, review websites and news articles, and surveys, provides essential insights for developing an effective growth strategy, which is essential for the long-term success of the business.

Reference

Aichner, T., Grünfelder, M., Maurer, O. and Jegeni, D., 2021. Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, behaviour, and social networking*, *24*(4), pp.215-222.

Aljedaani, W., Rustam, F., Mkaouer, M.W., Ghallab, A., Rupapara, V., Washington, P.B., Lee, E. and Ashraf, I., 2022. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, *255*, p.109780.

Busalim, A.H. and Ghabban, F., 2021. Customer engagement behaviour on social commerce platforms: An empirical study. *Technology in Society*, *64*, p.101437.

Chatterjee, S. and Kar, A.K., 2020. Why do small and medium enterprises use social media marketing and what is the impact: Empirical insights from India. *International Journal of Information Management*, *53*, p.102103.

Cheung, M.L., Pires, G.D. and Rosenberger III, P.J., 2019. Developing a conceptual model for examining social media marketing effects on brand awareness and brand image. *International Journal of Economics and Business Research*, *17*(3), pp.243-261.

Coates, A.E., Hardman, C.A., Halford, J.C., Christiansen, P. and Boyland, E.J., 2019. Social media influencer marketing and children's food intake: a randomized trial. *Pediatrics*, *143*(4).

Coco, S.L. and Eckert, S., 2020. # sponsored: Consumer insights on social media influencer marketing. *Public Relations Inquiry*, *9*(2), pp.177-194.

Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R. and Lozano-Quilis, J.A., 2020. Customer relationship management: digital transformation and sustainable business model innovation. *Economic research-Ekonomska istraživanja*, *33*(1), pp.2733-2750.

Gómez, M., Lopez, C. and Molina, A., 2019. An integrated model of social media brand engagement. *Computers in Human Behavior*, *96*, pp.196-206.

Gupta, S., Nawaz, N., Alfalah, A.A., Naveed, R.T., Muneer, S. and Ahmad, N., 2021. The relationship of CSR communication on social media with consumer purchase intention and brand admiration. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), pp.1217-1230.

Habel, J., Kassemeier, R., Alavi, S., Haaf, P., Schmitz, C. and Wieseke, J., 2020. When do customers perceive customer centricity? The role of a firm's and salespeople's customer orientation. *Journal of Personal Selling & Sales Management*, *40*(1), pp.25-42.

Hajba, G.L., 2018. Website Scraping with Python. *Berkeley: Apress*.

Hu, M. and Chaudhry, S.S., 2020. Enhancing consumer engagement in e-commerce live streaming via relational bonds. *Internet Research*, *30*(3), pp.1019-1041.

Islam, J.U., Hollebeek, L.D., Rahman, Z., Khan, I. and Rasool, A., 2019. Customer engagement in the service context: An empirical investigation of the construct, its antecedents and consequences. *Journal of Retailing and Consumer Services*, *50*, pp.277-285.

Jin, S.V., Muqaddam, A. and Ryu, E., 2019. Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, *37*(5), pp.567-579.

Krishen, A.S., Dwivedi, Y.K., Bindu, N. and Kumar, K.S., 2021. A broad overview of interactive digital marketing: A bibliometric network analysis. *Journal of Business Research*, *131*, pp.183-195.

Lawes, R., 2023. *Using semiotics in marketing: How to achieve consumer insight for brand growth and profits*. Kogan Page Publishers.

Lawson, R., 2015. *Web scraping with Python*. Packt Publishing Ltd.

Lee, S.M. and Lee, D., 2020. "Untact": a new customer service strategy in the digital age. *Service Business*, *14*(1), pp.1-22.

Li, F., Larimo, J. and Leonidou, L.C., 2021. Social media marketing strategy: definition, conceptualization, taxonomy, validation, and future agenda. *Journal of the Academy of Marketing Science*, *49*, pp.51-70.

Littlechild, S., 2021. Exploring customer satisfaction in Great Britain's retail energy sector Part I: The comparative use of Trustpilot online reviews in four sectors. Utilities Policy, 73, p.101298.

Mason, A.N., Narcum, J. and Mason, K., 2021. Social media marketing gained importance after Covid-19. *Cogent Business & Management*, *8*(1), p.1870797.

Meire, M., Hewett, K., Ballings, M., Kumar, V. and Van den Poel, D., 2019. The role of marketer-generated content in customer engagement marketing. *Journal of Marketing*, *83*(6), pp.21-42.

Mills, A.J. and Hair Jr, J.F., 2021. Consumer insights: a turning point for marketing research education. *Journal of Marketing Education*, *43*(3), pp.279-284.

Mitchell, R., 2018. *Web scraping with Python: Collecting more data from the modern web*. " O'Reilly Media, Inc.".

Moriuchi, E., 2019. Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*, *36*(5), pp.489-501.

Ouyang, X., Zhou, P., Li, C.H. and Liu, L., 2015, October. Sentiment analysis using convolutional neural network. In *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing* (pp. 2359-2364). IEEE.

Rani, S. and Kumar, P., 2019. Deep learning based sentiment analysis using convolution neural network. *Arabian Journal for Science and Engineering*, *44*, pp.3305-3314.

Sashi, C.M., Brynildsen, G. and Bilgihan, A., 2019. Social media, customer engagement and advocacy: An empirical investigation using Twitter data for quick service restaurants. *International Journal of Contemporary Hospitality Management*, *31*(3), pp.1247-1272.

Sima, V., Gheorghe, I.G., Subić, J. and Nancu, D., 2020. Influences of the Industry 4.0 revolution on the human capital development and consumer behaviour: A systematic review. *Sustainability*, *12*(10), p.4035.

Syaifullah, J., Syaifudin, M., Sukendar, M.U. and Junaedi, J., 2021. Social media marketing and business performance of MSMEs during the COVID-19 pandemic. *The Journal of Asian Finance, Economics and Business*, *8*(2), pp.523-531.

Talahaturuson, E., Gumelar, A.B., Sooai, A.G., Sueb, S., Suprihatien, S., Altway, H.A., Lestari, C.S.N., Darmajanti, P., Fanani, U.Z., Hariyanti, T. and Nilowardono, S., 2022, September. Exploring Indonesian Netizen's Emotional Behavior Through Investment Sentiment Analysis Using TextBlob-NLTK (Natural Language Toolkit). In *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 244-248). IEEE.

Wang, J., Yu, L.C., Lai, K.R. and Zhang, X., 2016, August. Dimensional sentiment analysis using a regional CNN-LSTM model. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 225-230).

Wankhade, M., Rao, A.C.S. and Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), pp.5731-5780.

Yan, Y., Yang, H. and Wang, H.M., 2017, July. Two simple and effective ensemble classifiers for twitter sentiment analysis. In *2017 Computing Conference* (pp. 1386-1393). IEEE.

Yasa, N.N.K., Adnyani, I.G.A.D. and Rahmayanti, P.L.D., 2020. The influence of social media usage on the perceived business value and its impact on the business performance of Silver Craft Smes in Celuk Village, Gianyar-Bali. *Academy of Strategic Management Journal*, *19*(1), pp.1-10.

Ye, G., Hudders, L., De Jans, S. and De Veirman, M., 2021. The value of influencer marketing for business: A bibliometric analysis and managerial implications. *Journal of Advertising*, *50*(2), pp.160-

Appendix A: Research Project Plan

- **INTRODUCTION AND JUSTIFICATION**

The landscape of business and consumer interactions has undergone a significant transformation in the modern digital era due to social media's pervasive influence. An unprecedented reliance on social media platforms for marketing, communication, and brand promotion has resulted in businesses of all sizes quickly expanding into the digital space. This change has been brought about by the realisation that social media provides a direct channel to a huge and diverse audience, enabling businesses to engage with customers on an unheard-of scale (Nemes and Kiss, 2021). The use of artificial intelligence (AI) and machine learning (ML) techniques, such as sentiment analysis, has emerged as a crucial tool in this context for comprehending and foreseeing the effects of business activities on social media.

In recent years, sentiment analysis has gained popularity among organizations, companies, and governments in addition to scholars. The internet has become the major source of information for the general population as a result of its rising popularity. Users use a variety of internet sites to share their ideas and viewpoints. Platforms for social interaction, opinion sharing, and sentiment expression have become essential elements of people's daily livesHere's where AI and ML come into play, completely changing how companies use these platforms' power (Aljedaani et al. 2022).

The main goal of this study is to use sentiment analysis to explore the complex interactions that exist between businesses and social media. This method of analysis makes use of AI and ML to evaluate the feelings, attitudes, and opinions contained within textual data, providing a window into the general sentiment of consumers (Chauhan et al., 2021). Businesses can better understand how the public views their actions, goods, and services by analysing sentiment.

The enhancement of small business-oriented companies has been designated as the focal point of this study. Enterprises like "Bettys," "The Yorkshire Soap Company," "Black

Yak," and "Masons Yorkshire Gin" epitomize the vitality of small businesses that have harnessed the potential of social media to propel their expansion. These firms often leverage customer feedback to refine their offerings and enhance overall customer satisfaction. Nevertheless, the manual collection and analysis of this feedback can prove to be arduous and time-intensive. In this context, the utilization of AI-driven sentiment analysis emerges as a transformative solution with the potential to revolutionize the process.

Sentiment analysis is the process of gathering and examining people's thoughts, beliefs, and perceptions on a range of issues, products, and services. The volume of comments and evaluations about routine activities has dramatically increased on social media platforms, blogs, and other Internet-based apps. For businesses, governments, or individuals to acquire information and make wise judgments, the views of people can be valuable. However, there are plenty of challenges with sentiment analysis and the rating process. These challenges make it difficult to accurately evaluate sentiment and establish the proper sentiment polarity (positive or negative). Sentiment analysis uses text mining and natural language processing (NLP) to extract and evaluate subjective information from text (Nemes and Kiss, 2021).

As social media sites have grown in popularity, several other disciplines have emerged that specialize in analyzing them and their content to obtain the data they want. Sentiment analysis focuses on determining readers' attitudes toward a text by examining its content. Since it falls under the umbrella of Natural Language Processing (NLP), a lot of early research must have been done in this area. (Basarslan and Kayaalp, 2020).

The proposed study, which examines how social media affects businesses through the lenses of sentiment analysis and artificial intelligence, is extremely important in the context of modern business. The results of this study have the potential to have a significant impact on strategies, choices, and customer satisfaction as businesses around the world navigate the digital space to connect with their target audience. The importance of the study stems from its potential to transform business practices, strengthen small businesses, and advance the discussion on the symbiotic relationship between technology and successful business outcomes (Krishen et al. 2021).

- **RESEARCH QUESTION**

1. How does the web scraper extract the web contents from social websites like review websites or some blogging-type websites?
2. How the extracted contents are processed by the method of regular expression for picking up idle sentences without any emoji or any symbolic expressions and how the Python AI scratch is used to serve as a tool for implementing sentence cleaning, tokenization, etc.
3. How are the sentiments analysed through the pattern and TextBlob, which are mostly used as Python sentiment tools?
4. How are sentiments helping to improve customer satisfaction with the company's activities?

- **OBJECTIVES**

1. To develop a methodical approach to web content extraction from social websites, such as blogging and review platforms, using web scraping techniques to ensure the collection of pertinent data for analysis.
2. To create a thorough framework for data processing that uses regular expression techniques to efficiently recognise and separate meaningful sentences from the extracted content, Investigate the use of Python AI scripts to simplify tokenization, sentence cleaning, and other crucial preprocessing steps.
3. To analyse and interpret the sentiments embedded in the textual data, use established Python sentiment analysis tools like Pattern and TextBlob. Examine how these tools can be used to correctly categorise feelings as positive, negative, or neutral.
4. To consider how the results of sentiment analysis can be used in practice to raise client satisfaction with a business's operations. Examine the ways in which sentiment insights can aid in the development of customer-centric strategies and serve as a strategic decision-making tool.

- **DELIVERABLE**

The analysis of effects is especially important for small businesses because they frequently have limited resources and rely heavily on customer loyalty. Companies like Bettys, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin are the best examples of this demographic, and it is strategically important to comprehend how they interact online (Khan et al. 2020). These companies can improve their marketing tactics, boost customer satisfaction, and spur growth by revealing sentiment patterns in their interactions.

Deliverable Items – Word Documents or PDF file, Technical File i.e. in GOOGLE COLAB format, survey consent files in SPSS format and PPT for presentation

- **LITERATURE REVIEW**

The literature review synthesises existing research on the relationship between social media, artificial intelligence (AI), and business outcomes in a critical manner. It highlights the importance of sentiment analysis in understanding consumer attitudes and behaviours. The review highlights the study's relevance by focusing on small businesses and their potential for expansion using sentiment-driven tactics.

- **VARIABLE**

1. **Making Strategic Decisions:** The foundation of the research's justification is the idea that a well-informed strategy complements consumer sentiment. Businesses can make decisions that resonate with their target audience by deciphering the undercurrents of sentiment in consumer discussions.
2. **Enhanced Customer Satisfaction:** By exposing the feelings that surround business activities, one can better predict what customers will expect. This encourages brand loyalty and repeat business, which are crucial for small businesses looking to grow sustainably.
3. **Data-Driven Insights:** By combining sentiment analysis and AI, data-driven insights are provided to businesses. These insights go beyond intuition

and speculative opinion, empowering businesses to make wise decisions and stay clear of potential pitfalls.

4. **Competitive Advantage:** Companies that use sentiment analysis have an advantage over their competitors. They are able to foresee changes in consumer preferences and adjust their offerings accordingly, outperforming their rivals.

- **IMPLEMENTATION**

The purpose of this project is to look into the potential effects of social media on commercial enterprises; especially those geared towards small businesses like Betty's, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin. The research will make use of machine learning (ML) and artificial intelligence (AI) methodologies to accomplish this. Whole technical was implemented on google colab platform. And coding language was chosen in python.

- **RESEARCH DESIGN**

Step 1: Retrive the comments from the review websites or blogging websites by using web scraping tools in Python.

Step 2: After web scraping, perform sentence cleaning, tokenization, and visualising the word cloud by word frequencies.

Step 3: Sentiment analysis by combined and average methods through the pattern and Textblob with NLTK (Natural Language Toolkit, which is NLP, i.e., Natural Language Process-oriented).

Step 4: Creating a dataset table saved in the format of CSV for classifying the sentences into two parts, i.e., subjectivity and polarity It is briefly said that one's opinion focuses on his or her comment’s remarkable sentiment or emotion. Also show the scatter characteristics between two variables, i.e., subjectivity and polarity, which are two major characteristics of the TextBlob and Pattern Python libraries.

Step 5: Separately identify the positive and negative words that can impact small businesses.

Step 6: Visualise the word cloud based on positive and negative words.

Step 7: Find the sentences that contain those positive and negative words according to the sorting for identifying the key terms for the requirements of the company's improvements.

Step 8: Also apply steps 1 to 7 for the other four companies, and the whole result is combined to be stored as a grand csv file for classification.

Step 9: However, the grand dataset comes from the primary analysis taken by textblob and pattern, and after this dataset is post-classified and fed onto the CNN layers,

Step 10: Final conclusion with remarks is to be made on the basis of Sentiment score analysis.

- **ETHICS RISKS AND ISSUES**

The profound influence of social media on the rapidly changing business landscape has completely changed how businesses interact with their target markets and promote their goods and services. Businesses face the difficult challenge of comprehending and anticipating the complex effects of their activities in this digital space as they increasingly recognise the potential of social media platforms as tools for brand exposure, customer interaction, and growth (Piedrahita-Valdes et al. 2021). Deciphering the complex relationships between businesses' social media activities and the opinions of the customers they serve is the main issue this study attempts to solve.

The urgent issue in this situation is best stated as follows: How can companies use social media platforms to effectively promote their products while also understanding the feelings, viewpoints, and reactions of their customer base? For small business-oriented organisations like "Bettys," "The Yorkshire Soap Company," "Black Yak," and

"Masons Yorkshire Gin," where growth and success heavily depend on forging a resonance with their clientele in the virtual space, this challenge becomes particularly pertinent.

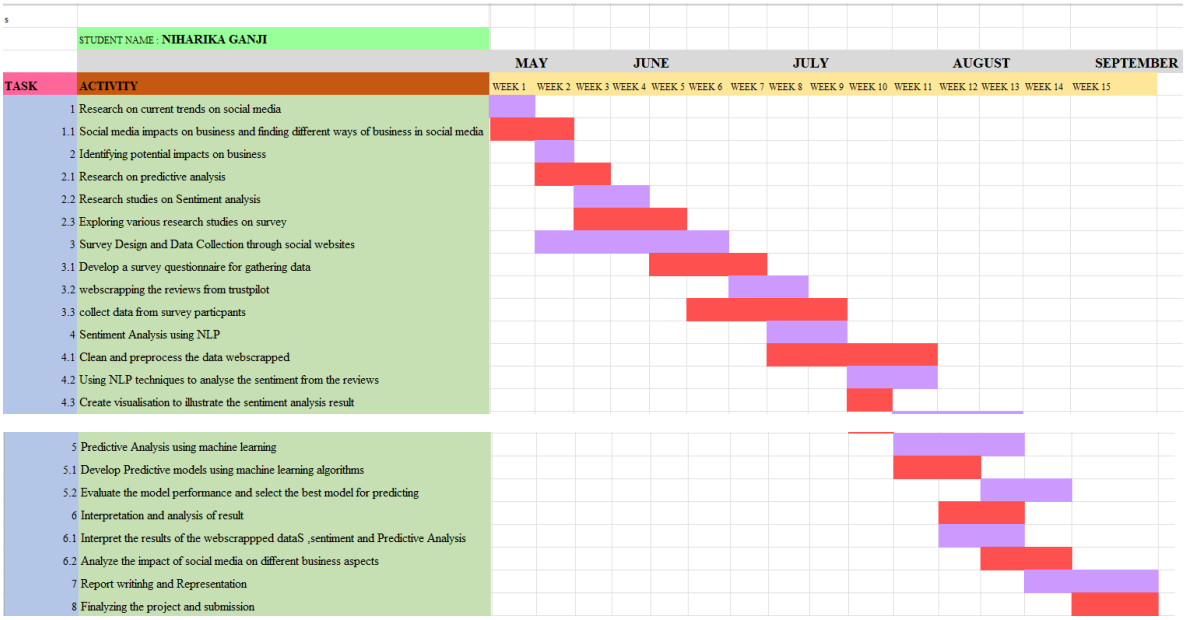
Due to the overwhelming volume of unstructured textual data generated by the social media industry's explosive growth, it is now nearly impossible for businesses to manually process and thoroughly analyse consumer sentiment. Furthermore, it can be difficult to distinguish between flattering compliments and sincere expressions of satisfaction. In order to overcome these difficulties, this research employs AI and ML techniques, particularly sentiment analysis, as a means of extracting insightful data from the substantial digital footprint that users leave across various online platforms (Aljedaani et al. 2022).

The choice of appropriate sentiment analysis tools and methodologies is also a problem. Finding the most precise and trustworthy options that fit the needs of small businesses can be difficult given the variety of tools at hand. Making educated choices about the tools that will accurately interpret and categorise the range of emotions expressed in customer reviews, comments, and interactions is the difficult part.

Additionally, there are dangers introduced by sentiment analysis's inherent subjectivity. Complex expressions must be labelled as positive, negative, or neutral with a nuanced understanding of language and context (Asif et al. 2020). Misinterpretations or incorrect classifications may produce flawed insights that misdirect business strategies, resulting in opportunities lost or efforts put in the wrong direction.

In order to accurately predict and understand the effects of businesses on social media, a comprehensive approach that combines AI, ML, and sentiment analysis is required. By enabling small business-oriented companies to make data-driven decisions that are in line with the preferences and expectations of their online audience, this research aims to close the gap between businesses' digital initiatives and consumer sentiments.

- **TIME PLAN**



STUDENT NAME : **NIHARIKA GANJI**

		MAY				JUNE				JULY				AUGUST				SEPTEMBER
TASK	ACTIVITY	WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 5	WEEK 6	WEEK 7	WEEK 8	WEEK 9	WEEK 10	WEEK 11	WEEK 12	WEEK 13	WEEK 14	WEEK 15		
1	Research on current trends on social media																	
1.1	Social media impacts on business and finding different ways of business in social media																	
2	Identifying potential impacts on business																	
2.1	Research on predictive analysis																	
2.2	Research studies on Sentiment analysis																	
2.3	Exploring various research studies on survey																	
3	Survey Design and Data Collection through social websites																	
3.1	Develop a survey questionnaire for gathering data																	
3.2	webscrapping the reviews from trustpilot																	
3.3	collect data from survey particpants																	
4	Sentiment Analysis using NLP																	
4.1	Clean and preprocess the data webscrapped																	
4.2	Using NLP techniques to analyse the sentiment from the reviews																	
4.3	Create visualisation to illustrate the sentiment analysis result																	
5	Predictive Analysis using machine learning																	
5.1	Develop Predictive models using machine learning algorithms																	
5.2	Evaluate the model performance and select the best model for predicting																	
6	Interpretation and analysis of result																	
6.1	Interpret the results of the webscrappped dataS ,sentiment and Predictive Analysis																	
6.2	Analyze the impact of social media on different business aspects																	
7	Report writinhg and Representation																	
8	Finalyzing the project and submission																	

Appendix B: Ethics form

UREC2 RESEARCH ETHICS PROFORMA FOR STUDENTS UNDERTAKING LOW RISK PROJECTS WITH HUMAN PARTICIPANTS

This form is designed to help students and their supervisors to complete an ethical scrutiny of proposed research. The University Research Ethics Policy should be consulted before completing the form. The initial questions are there to check that completion of the UREC 2 is appropriate for this study. The final responsibility for ensuring that ethical research practices are followed rests with the supervisor for student research.

Note that students and staff are responsible for making suitable arrangements to ensure compliance with the General Data Protection Act (GDPR). This involves informing participants about the legal basis for the research, including a link to the University research data privacy statement and providing details of who to complain to if participants have issues about how their data was handled or how they were treated (full details in module handbooks). In addition the act requires data to be kept securely and the identity of participants to be anonymized. They are also responsible for following SHU guidelines about data encryption and research data management. Information on the Ethics Website

The form also enables the University and College to keep a record confirming that research conducted has been subjected to ethical scrutiny.

The form may be completed by the student and the supervisor and/or module leader (as applicable). In all cases, it should be counter-signed by the supervisor and/or module leader, and kept as a record showing that ethical scrutiny has occurred. Some courses may require additional scrutiny. Students should retain a copy for inclusion in their research projects, and a copy should be uploaded to the relevant module Blackboard site.

Please note that it may be necessary to conduct a health and safety risk assessment for the proposed research. Further information can be obtained from the College Health and Safety Service.

Checklist Questions to ensure that this is the correct form

1. Health Related Research with the NHS or Her Majesty's Prison and Probation Service (HMPPS)or with participants unable to provide informed consent

Question	**Yes/No**
1. Does the research involve? • Patients recruited because of their past or present use of the NHS	NO
• Relatives/carers of patients recruited because of their past or present use of the NHS	NO
• Access to data, organs or other bodily material of past or present NHS patients	NO
• Foetal material and IVF involving NHS patients	NO
• The recently dead in NHS premises	NO
• Prisoners or others within the criminal justice system recruited for health-related research*****	NO
• Police, court officials, prisoners or others within the criminal justice system*****	NO
• Participants who are unable to provide informed consent due to their incapacity even if the project is not health related	NO
2. Is this a research project as opposed to service evaluation or audit? *For NHS definitions of research etc. please see the following website* http://www.hra.nhs.uk/documents/2013/09/defining-research.pdf	NO

If you have answered **YES** to questions **1 & 2** then you **MUST** seek the

appropriate external approvals from the NHS, Her Majesty's Prison and Probation Service (HMPPS) under their independent Research Governance schemes. Further information is provided below.
https://www.myresearchproject.org.uk

NB College Teaching Programme Research Ethics Committees (CTPRECS) provide Independent Scientific Review for NHS or HMPPS research and initial scrutiny for ethics applications as required for university sponsorship of the research. Applicants can use the IRAS proforma and submit this initially to their CTPREC.

1. **Checks for Research with Human Participants**

Question	**Yes/No**
1. Will any of the participants be vulnerable? *Note: Vulnerable' people include children and young people, people with learning disabilities, people who may be limited by age or sickness, people researched*	NO
2. Are drugs, placebos or other substances (e.g. food substances, vitamins) to be administered to the study participants or will the study involve	NO
3. Will tissue samples (including blood) be obtained from participants?	NO
4. Is pain or more than mild discomfort likely to result from the study?	NO
5. Will the study involve prolonged or repetitive testing?	NO
6. Is there any reasonable and foreseeable risk of physical or emotional harm to any of the participants? *Note: Harm may be caused by distressing or intrusive interview questions,*	NO
7. Will anyone be taking part without giving their informed consent?	NO
8. Is it covert research? *Note: 'Covert research' refers to research that is conducted without the knowledge of*	NO
9. Will the research output allow identification of any individual who	NO

If you have answered **YES** to any of these questions you are **REQUIRED** to complete and

submit a UREC 3 or UREC4). Your supervisor will advise. If you have answered **NO** to all these questions then proceed with this form (UREC 2).

General Details

Name of student	Niharika Ganji
SHU email address	Niharika.Ganji@student.shu.ac.uk
Course or qualification (student)	MSc in Artificial Intelligence
Name of supervisor	Muhammad Aitsam
email address	M.Aitsam@shu.ac.uk
Title of proposed research	Predicting the impacts of business on the social media
Proposed start date	22nd of may
Proposed end date	7th of sep
Background to the study and scientific rationale for undertaking it.	
Aims & research question(s)	The aim of this project is to predict the impacts of social media on business. The data is collected through social websites. In this dissertation, this project's goal is to investigate how social media may affect commercial operations, particularly those dedicated to small firms like Betty's, The Yorkshire Soap Company, Black Yak, and Masons Yorkshire Gin. To do this, the project will employ machine learning (ML) and artificial intelligence (AI) techniques.
Methods to be used for: 1. recruitment of participants, 2.data collection,	1. Through the survey participants give there views on the social media business 2. Data is collected through some social websites. 3.Sentiment Analysis

3. data analysis.	
Outline the nature of the data held, details of anonymisation, storage and disposal procedures as required.	Yes, storage of data is being maintained.

3. Research in Organisations

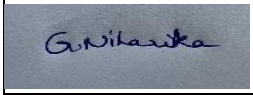
Question	**Yes/No**
1. Will the research involve working with/within an organisation (e.g. school, business, charity, museum, government department, international	NO
2. If you answered YES to question 1, do you have granted access to conduct the research? *If YES, students please show evidence to your supervisor. PI should retain safely.*	
3. If you answered NO to question 2, is it because: A. you have not yet asked B. you have asked and not yet received an answer	

4. Research with Products and Artefacts

Question	**Yes/No**
1. Will the research involve working with copyrighted documents, films, broadcasts, photographs, artworks, designs, products, programmes, databases,	Yes

2. If you answered YES to question 1, are the materials you intend to use in the public domain? *Notes: 'In the public domain' does not mean the same thing as 'publicly accessible'.* • *Information which is 'in the public domain' is no longer protected by copyright (i.e. copyright has either expired or been waived) and can be used without permission.* • *Information which is 'publicly accessible' (e.g. TV broadcasts, websites, artworks, newspapers) is available for anyone to consult/view. It is still protected by copyright even if there is no copyright notice. In UK law, copyright protection is automatic and does not require a copyright statement, although it*	Yes
3. If you answered NO to question 2, do you have explicit permission to use these materials as data? *If YES, please show evidence to your supervisor.*	
4. If you answered NO to question 3, is it because: A. you have not yet asked permission B. you have asked and not yet received and answer C. you have asked and been refused access.	**A/B/C**

Adherence to SHU policy and procedures

Personal statement	
I can confirm that: • I have read the Sheffield Hallam University Research Ethics Policy and Procedures	
Student	
Name: Niharika Ganji	Date: 01/08/2023
Signature: 	
Supervisor or other person giving ethical sign-off	

I can confirm that completion of this form has not identified the need for ethical approval by the FREC or an NHS, Social Care or other external REC. The research will not commence until any approvals required under Sections 3 & 4 have been received and any necessary health and safety measures are in place.	
Name: Muhammad Aitsam	Date: 01/08/2023
Signature: 	
Additional Signature if required by course:	
Name:	Date:
Signature:	

Please ensure the following are included with this form if applicable, tick box to indicate:

	Yes	**No**	**N/A**
Research proposal if prepared previously	☐✓	☐	☐
Any recruitment materials (e.g. posters, letters, etc.)	☐	☐✓	☐
Participant information sheet	☐✓	☐	☐
Participant consent form	☐✓	☐	☐
Details of measures to be used (e.g. questionnaires, etc.)	☐✓	☐	☐
Outline interview schedule / focus group schedule	☐	☐	☐✓
Debriefing materials	☐	☐	☐✓
Health and Safety Project Safety Plan for Procedures	☐	☐	☐✓

Appendix C:Information Sheet and Consent forms

PREDICTING THE IMPACT OF BUSINESS ON SOCIAL MEDIA

The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of **public tasks that are in the public interest.** A full statement of your rights can be found at: www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research. However, all University research is reviewed to ensure that participants are treated appropriately and their rights respected. This study was approved by UREC with Converis number ER*xxxxxxx*. Further information at: www.shu.ac.uk/research/excellence/ethics-and-integrity

1. Please, will you take part in this study as it is related to my study about social media usage for business purposes? How would you feel about future social media business, and what are the expectations? We want to know your opinion on business made through online social media apps.

2. You are free to choose whether or not you want to participate. If you choose to participate, you may keep a copy of the information supplied here and the consent form.

3. You can still opt not to answer a question or withdraw at any moment without providing a justification.

4. You need to answer simple questions that are being asked through survey questionnaires about the social media business and about the future implementation or expectations and drawbacks of the social media business.

5. The information you will provide will only be used to predict the future business value in social media and how the business people can improve to reach the expectation and get benefits and customer satisfaction.

6. This survey will only take place once. There are no risks or disadvantages in taking part in this survey. By taking this survey public can expect good service from online social media businesses in future.

7. There will be No one to contact you about the survey which has been taken about the social media business. It is just used for future improvement of business, benefits and future expectations of people.

8. Researchers are responsible for all the information that public participants provide.

9. The information the public provides through a survey can be accessed by the business person anonymously without any personal information.

10. The information is used for further studies after the study is over.

11. The information will be used for my project, report writing and presentation purposes. The study will last for 3 to 4 months.

12. After completing the study, you will be provided with a brief report on how your information has been used and helpful for the project.

If there are any questions about the survey, you can reach out to the researcher the contact details are given below.

Researcher Details:
Niharika Ganji
+447767946739

You should contact the Data Protection Officer if:	**You should contact the Head of Research Ethics (Dr Mayur Ranchordas) if:**
• you have a query about how your data is used by the University • you would like to report a data security breach (e.g. if you think your personal data has been lost or disclosed inappropriately) • you would like to complain about how the University has used your personal data DPO@shu.ac.uk	• you have concerns with how the research was undertaken or how you were treated ethicssupport@shu.ac.uk

Postal address: Sheffield Hallam University, Howard Street, Sheffield S1 1WBT

Telephone: 0114 225 5555

ADDING SOME OF THE PARTICIPANTS CONSENT FORMS

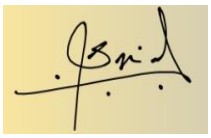
1ST PARTICIPANT CONSENT FORM

TITLE OF RESEARCH STUDY:PREDICTING THE IMPACT OF BUSINESS ON THE SOCI MEDIA

Please answer the following questions by ticking the response that applies

	YES	NO
1. I have read the Information Sheet for this study and have had details of the study explained to me.	✓☐	☐
2. My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point.	✓☐	☐
3. I understand that I am free to withdraw from the study within the time limits outlined in the Information Sheet, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher.	✓☐	☐
4. I agree to provide information to the researchers under the conditions of confidentiality set out in the Information Sheet.	✓☐	☐
5. I wish to participate in the study under the conditions set out in the Information Sheet.	✓☐	☐
6. I consent to the information collected for the purposes of this	✓☐	☐

research study, once anonymised (so that I cannot be identified), to be used for any other research purposes.



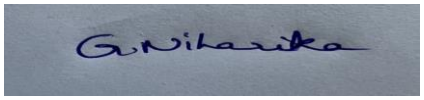
Participant's Signature:

Date: 03/05/2023

Participant's Name (Printed): Sujeevan Pamula

Contact details: Sujeevan,Eastbank road, 52myrtle close,S2 3HW,+44 7776817274

Researcher's Name (Printed): Niharika Ganji

Researcher's Signature: 

Researcher's contact details:

Niharika Ganji

18leadmill road, archways, S1 4SG

+44 7767946739

Please keep your copy of the consent form and the information sheet together.

2ND PARTICIPANT CONSENT FORM

TITLE OF RESEARCH STUDY: PREDICTING THE IMPACT OF BUSINESS ON THE SOCIAL MEDIA

Please answer the following questions by ticking the response that applies

	YES	NO
1.I have read the Information Sheet for this study and have had details of the study explained to me.	☐✓	☐
2.My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point.	☐✓	☐
3.I understand that I am free to withdraw from the study within the time limits outlined in the Information Sheet, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher.	☐✓	☐
4.I agree to provide information to the researchers under the conditions of confidentiality set out in the Information Sheet.	☐✓	☐
5.I wish to participate in the study under the conditions set out in the Information Sheet.	☐✓	☐
6.I consent to the information collected for the purposes of this research study, once anonymised (so that I cannot be identified), to be used for any other research purposes.	☐✓	☐

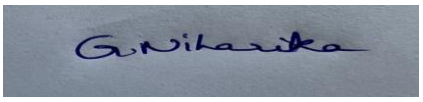
Participant's Signature: 

Date: 03/05/2023

Participant's Name (Printed): Srikanth

Contact details: Srikanth,52 myrtle close,S2 3HW+44 7760465679

Researcher's Name (Printed): Niharika Ganji

Researcher's Signature: 

Researcher's contact details:

Niharika Ganji

18leadmill road, archways, S1 4SG

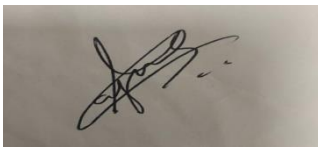
+44 7767946739

3RD PARTICIPANT CONSENT FORM

TITLE OF RESEARCH STUDY: PREDICTING THE IMPACT OF BUSINESS ON THE SOCIAL MEDIA

Please answer the following questions by ticking the response that applies

	YES	NO
1.I have read the Information Sheet for this study and have had details of the study explained to me.	☐✓	☐
2.My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point.	☐✓	☐
3.I understand that I am free to withdraw from the study within the time limits outlined in the Information Sheet, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher.	☐✓	☐
4.I agree to provide information to the researchers under the conditions of confidentiality set out in the Information Sheet.	☐✓	☐
5.I wish to participate in the study under the conditions set out in the Information Sheet.	☐✓	☐
6.I consent to the information collected for the purposes of this research study, once anonymised (so that I cannot be identified), to be used for any other research purposes.	☐✓	☐

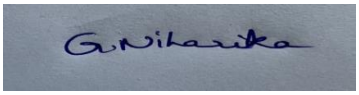
Participant's Signature: 

Date: 03/05/2023

Participant's Name (Printed): Arjun Sunil Kumar

Contact details: Arjun,53 myrtle close, S2 3HW,+44 7836275144

Researcher's Name (Printed): Niharika Ganji

Researcher's Signature: 

Researcher's contact details:

Niharika Ganji

18leadmill road, archways, S1 4SG

+44 7767946739

Appendix D: Data Gathering, Survey Analysis

- **Survey Questions:**

1. How frequently do you interact with companies on social mediA, on a scale of 1 to 5?

2. What kind of information or conversations do you value most when interacting with brands on social media?

3. How satisfied are you generally with the way businesses interact with their consumers on social media?

4. Can you think of a recent engagement you had with a company on social media? If yes, could you briefly describe it?

5. What changes or adjustments would you advise businesses to boost their social media presence and client interactions?

6. Do you think that people's reviews from Trustpilot websites might be quite similar as compared to Facebook, twitter and Instagram, like popular social websites?

7.How much do you trust the information and ads posted by companies on social media, on a scale of 1 to 5, with one being strongly disagree and five being strongly agree?

8. What do you think of the customised offers and information that you get from companies on social media platforms?

- **Demographic Profile:**

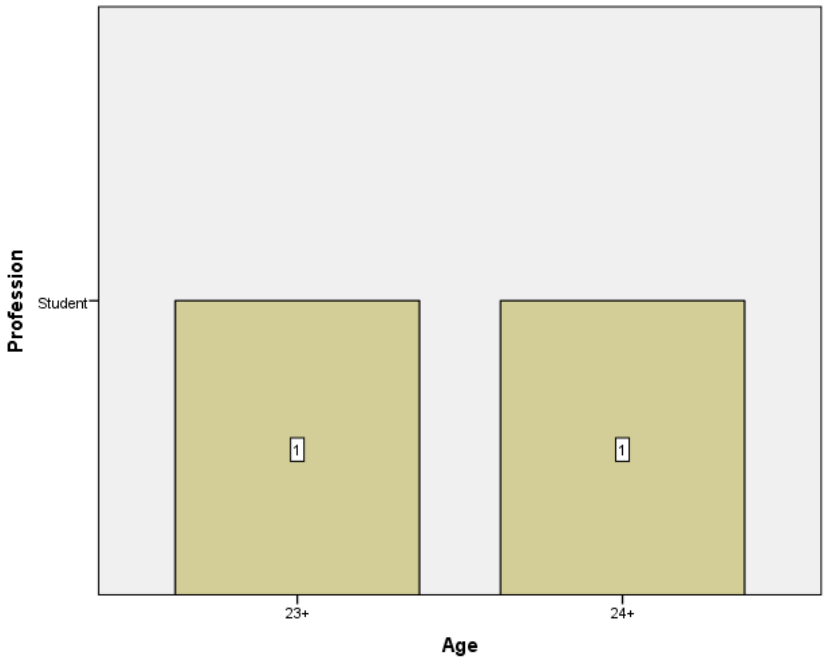


Fig: Demographic profile for a set of {Profession, Age}

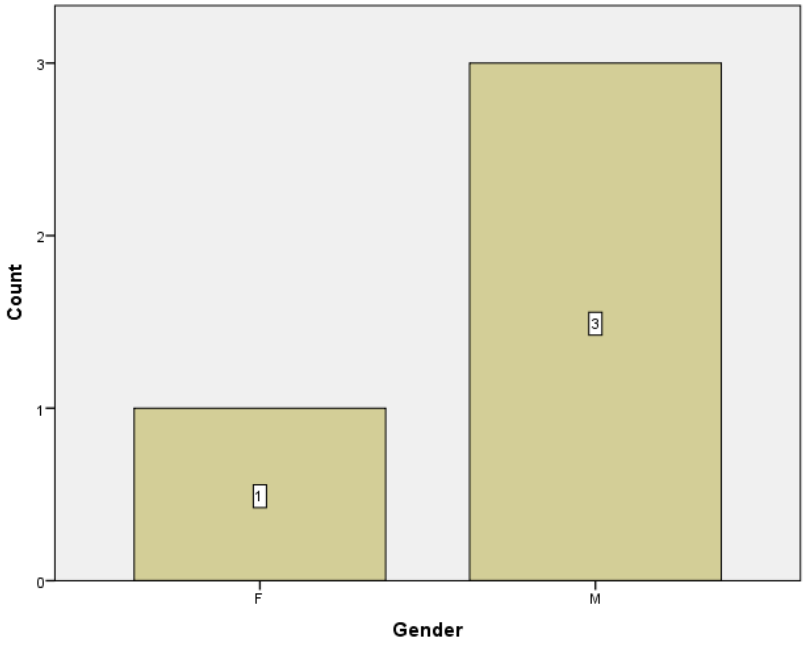


Fig: Demographic profile for a Gender

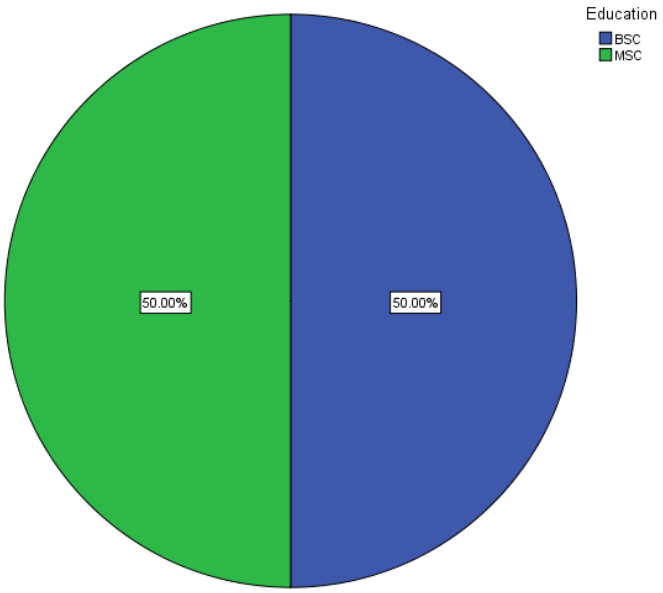


Fig: Demographic profile for Education

- **Conclusion over responses:**

Overall, the survey highlights the importance of organisations' social media presence and the necessity to continually improve approaches to increase client happiness and trust in this digital environment.