INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Analysis of consumer perception on Greenwashing through NLP: contributions to marketing strategy

Ricardo Filipe Salvador Pires de Melo

Master's in Business Analytics

Supervisor:

Professor Doctor Ricardo João Lourenço de Abreu,
Visiting Assistant Professor,
ISCTE-IUL

September, 2023

BUSINESS
SCHOOL

Department of Quantitative Methods for Management and Economics

Analysis of consumer perception on Greenwashing through NLP: contributions to marketing strategy

Ricardo Filipe Salvador Pires de Melo

Master's in Business Analytics

Supervisor:

Professor Doctor Ricardo João Lourenço de Abreu,
Visiting Assistant Professor,
ISCTE-IUL

September, 2023

This dissertation is dedicated to my loving family, supportive friends, and those I have crossed paths with on this academic journey. Your collective influence and support have shaped my approach and enriched this endeavor.

Acknowledgments

I want to take this opportunity to extend my sincere appreciation to the individuals who have been instrumental in completing my dissertation for my Master's in Business Analytics.

First and foremost, I owe a debt of gratitude to my dedicated supervisor, Prof. Ricardo. Your unwavering support, scholarly guidance, and insightful feedback have shaped this work and my growth as a researcher. Your mentorship has been invaluable.

I also want to express my gratitude to Prof. Raul, the director of the Business Analytics program, for providing a robust academic environment and opportunities for learning and growth. Your leadership has been inspiring.

To all my teachers throughout this journey, thank you for imparting knowledge, igniting my curiosity, and nurturing my passion for business analytics. Your collective wisdom has been the foundation of my academic achievements.

I sincerely thank my family and friends for standing by me throughout this challenging yet rewarding journey. Your unwavering support, encouragement, and belief in my abilities have motivated me, even in the most demanding times.

This dissertation is a testament to the collaborative effort of these remarkable individuals. I am profoundly grateful for their contributions to this significant milestone in my academic voyage.

Resumo

Esta dissertação aborda uma lacuna na literatura, analisando de forma abrangente as reações dos consumidores ao greenwashing. Utilizando o Processamento de Linguagem Natural, distingue reações positivas e negativas e identifica emoções como a alegria, a tristeza e o desgosto. Esta investigação lança luz sobre os sentimentos dos consumidores em relação ao greenwashing, oferecendo conhecimentos práticos às empresas para melhorar o marketing e evitar a perceção do greenwashing.

A revisão da literatura, orientada pela abordagem PRISMA, concentra-se no greenwashing, no comportamento do consumidor e na comunicação ambiental. A seleção da literatura obedeceu a critérios rigorosos, provenientes das bases de dados Web of Science e Scopus, conduzindo à análise e avaliação de 23 artigos científicos.

Seguindo a metodologia CRISP-DM, esta dissertação recolheu dados do Twitter, centrando-se em tweets de empresas suspeitas de greenwashing. Uma análise das respostas a estes tweets incluiu considerações como a indústria, a apresentação das alegações, o tipo de elogio e as ações mencionadas. Utilizando o modelo de linguagem BERT, as respostas foram categorizadas com base no sentimento e nas emoções. Foram efetuadas várias análises, incluindo avaliações bi-variadas, visualizações de nuvens de etiquetas, regressão logística e testes de qui-quadrado.

Os resultados indicam que os temas relacionados com o clima não constituíam a principal preocupação dos consumidores. O sentimento negativo esteve presente nas respostas, mas também foi expressa alegria. As ações substanciais geraram respostas mais positivas. Em conclusão, a ação substantiva e o elogio ao consumidor são mais eficazes para cultivar reações positivas e mitigar o greenwashing.

Palavras-chave: Revisão Sistemática da Literatura, Greenwashing, Marketing Verde, Perceção do Consumidor, Processamento de Linguagem Natural

JEL Sistema de Classificação: Marketing (M31); Responsabilidade Social (M14)

Abstract

This dissertation addresses a critical literature gap by comprehensively analyzing consumer responses to greenwashing. Utilizing Natural Language Processing (NLP) distinguishes positive and negative reactions and identifies emotions like joy, sadness, and disgust. This research sheds light on consumer sentiments toward greenwashing, offering actionable insights for businesses to enhance marketing and prevent greenwashing perception.

The literature review, guided by the PRISMA approach, concentrates on greenwashing, consumer behavior, and environmental communication. The selection of literature adhered to rigorous criteria sourced from Web of Science and Scopus databases, leading to the analysis and evaluation of 23 scientific articles.

Following the CRISP-DM methodology, this dissertation collected data from Twitter, focusing on corporate tweets suspected of greenwashing. A comprehensive examination of responses to these tweets included considerations such as industry sector, claim presentation, praise type, and mentioned actions. Leveraging the BERT language model, responses were categorized based on sentiment and emotions. Various analyses were conducted, encompassing bivariate assessments, tag cloud visualizations, logistic regression, and chi-square tests.

The results indicate that climate-related topics were not the primary concern for consumers. Negative sentiment was present in responses, but joy was also expressed. Compensation claims and net-zero terms had limited influence on sentiment. Substantive actions generated more positive responses. Some industry sectors, like Communication Services, Energy, Financial Services, and Industrial sectors, received notable negative responses. Corporate praise triggered stronger negative reactions than consumer praise. In conclusion, substantive action and consumer praise are more effective in cultivating positive reactions and mitigating greenwashing.

Keywords: Systematic Literature Review, Greenwashing, Green Marketing, Consumer Perception, Natural Language Processing

JEL Classification System: Marketing (M31); Social Responsibility (M14)

Table of Contents

Acknowledgments iii
Resumo v
Abstract vii
Table of Contents ix
List of figures xi
List of tables xii
List of acronyms xiii
Introduction 1
1.1. Introduction 1
1.2. Research questions 1
1.3. Objectives and validation forms 2
1.4. Research outline 3
Literature Review 5
2.1. Systematic Literature Review Protocol 5
2.1.1. Objectives and descriptive questions 5
2.1.2. Article Selection Process 6
2.1.3. Evaluation of articles from the systematic literature review 7
2.2. Summary of the articles' contents 8
2.3. Evaluation of articles from the systematic literature review 12
2.4. Greenwashing 14
2.5. Green marketing 15
2.5.1. Relevance of green marketing 15
2.5.2. CSR (Corporate Social Responsibility) 16
2.5.3. Greenwashing perception 17
2.5.4. Claims and communication 17
2.5.5. Greenwashing and emotions 22
2.6. Greenwashing and social media 23
2.7. Greenwashing and NLP 25
2.8. Answer to descriptive questions 26
Methodology 32

3.1. Business Understanding 32
3.2. Data Understanding 32
3.3. Data Preparation 33
3.3.1. Variable creation 33
3.3.2. Descriptive analysis 36
3.4. Modeling 43
3.5. Evaluation 44
3.6. Deployment 48
Results and Discussion 50
4.1. Consumer response analysis 50
4.2. Consumer sentiment analysis 55
4.3. Strategies for effective marketing campaigns 66
Conclusions 70
5.1. Conclusion 70
5.2. Limitations 72
Bibliography 74
Annexes 79
Annex A – Climate Dictionary 79
Annex B – Chi-square test 80

List of figures

Figure 1 - Article selection process. 7
Figure 2 - Data collection process 36
Figure 3 - Basic steps to perform sentiment analysis and emotion detection. Based on (Nandwani & Verma, 2021) 43
Figure 4 - Topic modeling LDA Replies 51
Figure 5 - TF-IDF Replies 52
Figure 6 - Bi-grams Replies 52
Figure 7 - Trigrams Replies 53
Figure 8 - Sector climate response 54
Figure 9 - Climate score response 54
Figure 10 - Sentiment distribution 55
Figure 11 - Emotion distribution 56
Figure 12 - Claim vs. Sentiment 57
Figure 13 - Claim vs. Emotion 57
Figure 14 - Action vs. Sentiment 58
Figure 15 - Action vs. Emotion 58
Figure 16 - Sector vs. Sentiment 59
Figure 17 - Sector vs. Emotion 60
Figure 18 - Praising vs. Sentiment 61
Figure 19 - Praising vs. Emotion 61
Figure 20 - Net-zero vs. Sentiment 62
Figure 21 - Net-zero vs. Emotion 62
Figure 22 - CDP Climate Change Score vs. Sentiment 63
Figure 23 - CDP Climate Change Score vs. Emotion 64

List of tables

Table 1 - Literature exclusion and inclusion criteria. 6
Table 2 - Criteria for evaluating articles from the systematic literature review. 8
Table 3 - Articles included in the systematic literature review. 8
Table 4 - Summary of the articles included in the literature review. 11
Table 5 - Evaluation of articles from the systematic literature review 13
Table 6 - Distribution of both corporate tweets and replies across the various sectors. 37
Table 7 - Summary of the distribution of corporate tweets and replies per company and sector. 38
Table 8 - Summary of the distribution of CDP scores across sectors. 39
Table 9 -Summary of the distribution of claims. 40
Table 10 - Summary of the distribution of praising. 41
Table 11 - Summary of the distribution of actions. 42
Table 12 - Summary of the distribution of net zero mentions. 42
Table 13 - Datasets for sentiment analysis and emotion detection. 45
Table 14 - Sentiment Analysis Models 46
Table 15 - Emotion Detection Models 48
Table 17 - Logistic regression model summary 65
Table 18 - Logistic regression classification table 65
Table 19 - Logistic regression classification metrics 65
Table 20 - Logistic regression variables in the equation 66

List of acronyms

API – Application Programming Interface

BERT - Bidirectional Encoder Representations with Transformers

CDP - Carbon Disclosure Project

CRISP-DM - Cross-Industry Standard Process for Data Mining

CSR - Corporate Social Responsibility

DNN - Deep Neural Network

FTC - Federal Trade Commission

GWWOM - Greenwashed Word-of-mouth

NGO – Non-governmental organization

NLP - Natural Language Processing

NLU - Natural Language Understanding

PRISMA - Preferred Reporting Items for Systematic reviews and Meta-Analyses

SOTA - State-of-the-Art

TCFD - Task Force on Climate-related Financial Disclosures

TF-IDF - Term Frequency-Inverse Document Frequency

WOM - Word-of-mouth

CHAPTER 1

Introduction

1.1. Introduction

Sustainability has gained significant attention recently, especially in the business and consumer spheres. Companies have embraced sustainable practices to appeal to consumers who are becoming increasingly environmentally conscious. However, not all sustainability claims are genuine, leading to the rise of "greenwashing," where companies make false or exaggerated environmental claims to promote their brand/product.

So-called "green" brands seek to promote their sustainability values through marketing campaigns to change the consumer's perception and increase sales. However, these marketing campaigns do not always convey the reality, consisting of false or incomplete information by companies. As described by Polonsky (1994), this phenomenon is called greenwashing. Social media platforms have also made it easier for consumers and activists to call attention to and mobilize opposition against companies that practice greenwashing notes (Lyon & Montgomery, 2013).

This research aims to fill a gap in the literature by conducting a broad analysis (without a specific focus on a particular company, event, or industry) of potential consumers' reactions to greenwashing. It intends to add a new dimension to previous studies by using NLP (natural language processing) techniques to differentiate between positive and negative responses and identify emotions such as sadness, joy, or disgust. By analyzing consumer opinions and sentiments towards greenwashing, this research aims to shed light on the issue and provide insights for companies to improve their marketing efforts and avoid the perception of greenwashing.

Additionally, the study is expected to develop a strategy to avoid the perception of greenwashing, considering the reactions and emotions of potential consumers. Using NLP techniques will help differentiate positive and negative responses and identify emotions, adding a new dimension to the previously conducted studies. The results of this research will provide valuable insights for companies to improve their marketing efforts and avoid the perception of greenwashing and miscommunication, leading to a more sustainable future for both businesses and the environment.

1.2. Research questions

This research analyzes consumer perception of greenwashing and its impact on marketing strategies. Three main research questions will be addressed:

1. The impact of green marketing campaigns: To determine the extent to which consumer responses to these campaigns are focused on greenwashing or other issues.
2. Consumer perceptions of green marketing campaigns: To investigate whether consumers' feelings about green marketing campaigns are influenced by the sector of the company running the campaign, as well as the content of the campaign, such as the claims made, the keywords used, and the actions mentioned.
3. Recommendations for effective green marketing campaigns: This research outlines recommendations for running marketing campaigns that are unlikely to be perceived as greenwashing. The research will identify common threads among campaigns that have met with little or no opposition and identify the factors that contribute to the opposition, wherever possible.

This study aims to provide valuable insights into consumer perception of greenwashing and contribute to developing effective marketing strategies for companies looking to communicate their sustainable efforts to consumers.

1.3. Objectives and validation forms

This study aims to thoroughly examine consumer reactions to green marketing campaigns to understand better the impact of false or incomplete information in these campaigns. The research seeks to achieve four main objectives using various validation forms and techniques.

The first objective is to evaluate the spectrum of consumer reactions to green marketing campaigns containing false or incomplete information. To validate this objective, the study will apply sentiment analysis techniques to data obtained through social media platforms. This data will be analyzed to gain insight into consumer reactions and feelings toward green marketing campaigns that contain false or incomplete information.

The second objective is to infer the relationship between consumer reactions and several distinct attributes, such as type of claims, actions, praising, company sector, and presence of specific keywords. To achieve this, the study will classify and compare the collected data to the present emotions and sentiments. This will shed light on the attributes most likely to elicit adverse consumer reactions to green marketing campaigns.

Finally, the third objective is to formulate a strategy to avoid the perception of greenwashing. Based on the previous points, the study will identify forms of communication least likely to elicit an adverse consumer reaction. This information will be used to develop a strategy companies can use to avoid the perception of greenwashing in their marketing efforts.

In conclusion, this study is designed to provide a comprehensive understanding of consumer reactions to green marketing campaigns and the impact of false or incomplete information in these campaigns. Using sentiment analysis and other techniques, the study aims to provide valuable insights for companies to improve their marketing efforts and avoid the perception of greenwashing.

1.4. Research outline

The proposed dissertation explores the topic of green marketing and its associated concepts, such as corporate social responsibility, greenwashing, green trust, and green word-of-mouth. The research seeks to answer several questions regarding the current state of green marketing and its impact on consumers.

Chapter 1 of the dissertation will introduce the topic of green marketing and describe the problem the research aims to address. It will also present the research questions and outline the structure of the thesis.

Chapter 2 will undertake a systematic literature review of the existing body of knowledge in green marketing, greenwashing, and consumer behavior. This review will explore concepts such as corporate social responsibility, green marketing, greenwashing, green trust, and green word-of-mouth and analyze how they have been studied.

Chapter 3 of the dissertation will describe the methodology used in the research, which will follow the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. This chapter will also outline the research approach and data collection methods.

Chapter 4 will present the research results, along with a discussion. The chapter will provide insights into the current state of green marketing and its impact on consumers.

Chapter 5 will summarize the research findings and outline the study's limitations.

Overall, this dissertation aims to comprehensively examine the topic of green marketing and its associated concepts. The research will contribute to the existing body of knowledge and inform future studies on the subject.

CHAPTER 2

Literature Review

2.1. Systematic Literature Review Protocol

This chapter presents a systematic literature review protocol for this research. A systematic literature review is a rigorous and systematic process that involves searching, evaluating, and synthesizing existing research studies related to a particular research question or topic. The protocol outlined in this chapter is based on the PRISMA methodology. This process, proposed by Kitchenham (2004), is designed to identify, evaluate, and synthesize the literature published by researchers in the area under study. The systematic literature review protocol outlines the methodology that will used to conduct the review, including the initial objectives and questions, the search and collection of literature, the inclusion and exclusion criteria, the data extraction and analysis methods, and the assessment of the articles to identify which ones are most relevant to the topic under study. This protocol aims to ensure a transparent and reproducible review process, minimize bias and errors, and provide a comprehensive literature summary in the area under investigation. This chapter describes the systematic literature review protocol and how it will be executed to achieve the research objectives.

2.1.1. Objectives and descriptive questions

The main objective of the systematic literature review is to summarize the scientific literature on greenwashing, its relationship with marketing strategies, and consumer perception. The review objectives are to contribute to the collection of knowledge on greenwashing, with a particular emphasis on two aspects: (1) analysis of consumer perception of green marketing campaigns accused of greenwashing and (2) recommendations for conducting a green marketing campaign.

The general descriptive question is, "What research has been conducted on evaluating consumer perceptions regarding green marketing campaigns and investigating potential instances of greenwashing?"

The specific descriptive questions are:

i. “How was the investigation into consumer perception of green marketing campaigns accused of greenwashing conducted?”

ii. “What factors should be considered when determining consumer perception of green marketing campaigns accused of greenwashing?”

iii. “How were the green marketing campaigns accused of greenwashing identified/constructed?”
iv. “What conclusions were drawn regarding consumer perception, and what factors influence perception?”
v. “What were the strategic recommendations for running a green marketing campaign?”

2.1.2. Article Selection Process

Scientific studies were automatically retrieved from scientific publication databases, Web of Science and Scopus, for review based on keywords, search period, language, and inclusion/exclusion criteria. These databases were selected based on their scope and wide range of publications in consumer behavior analysis and greenwashing (Mongeon & Paul-Hus, 2016).

An automatic search was performed on the selected sources, paying attention to keywords, periods, and language. The query used was *((“consumer” or "purchase behavior" or "purchase intention" or "theory of planned behavior") and ("environmental communication" or "environmental marketing" or "marketing" or "green advertising" or "green consumer" or "green marketing" or "green purchase decision" or "green purchasing intentions" or "green trust" or “csr” or “corporate social responsibility” or “green buying behavior”) and ("greenwash" or "greenwashing"))*.

The review process involved the application of inclusion and exclusion criteria, evaluation according to title and abstract, introduction and conclusion, and a complete reading of the article. The query application resulted in 125 articles from both databases, with 81 remaining after exclusion criteria, 63 after the title and abstract analysis, 40 after the introduction and conclusion analysis, and 23 after full-text analysis, which were included in the literature review. Table 1 presents the inclusion and exclusion criteria.

Table 1 - Literature exclusion and inclusion criteria.

Inclusion criteria
• Articles addressing greenwashing. • Articles addressing consumer behavior and perception. • Articles addressing environmental communication.
Exclusion criteria
• Articles outside the publication period 2018 to 2022. • Duplicated articles. • Articles published in languages other than English. • Articles without open access.

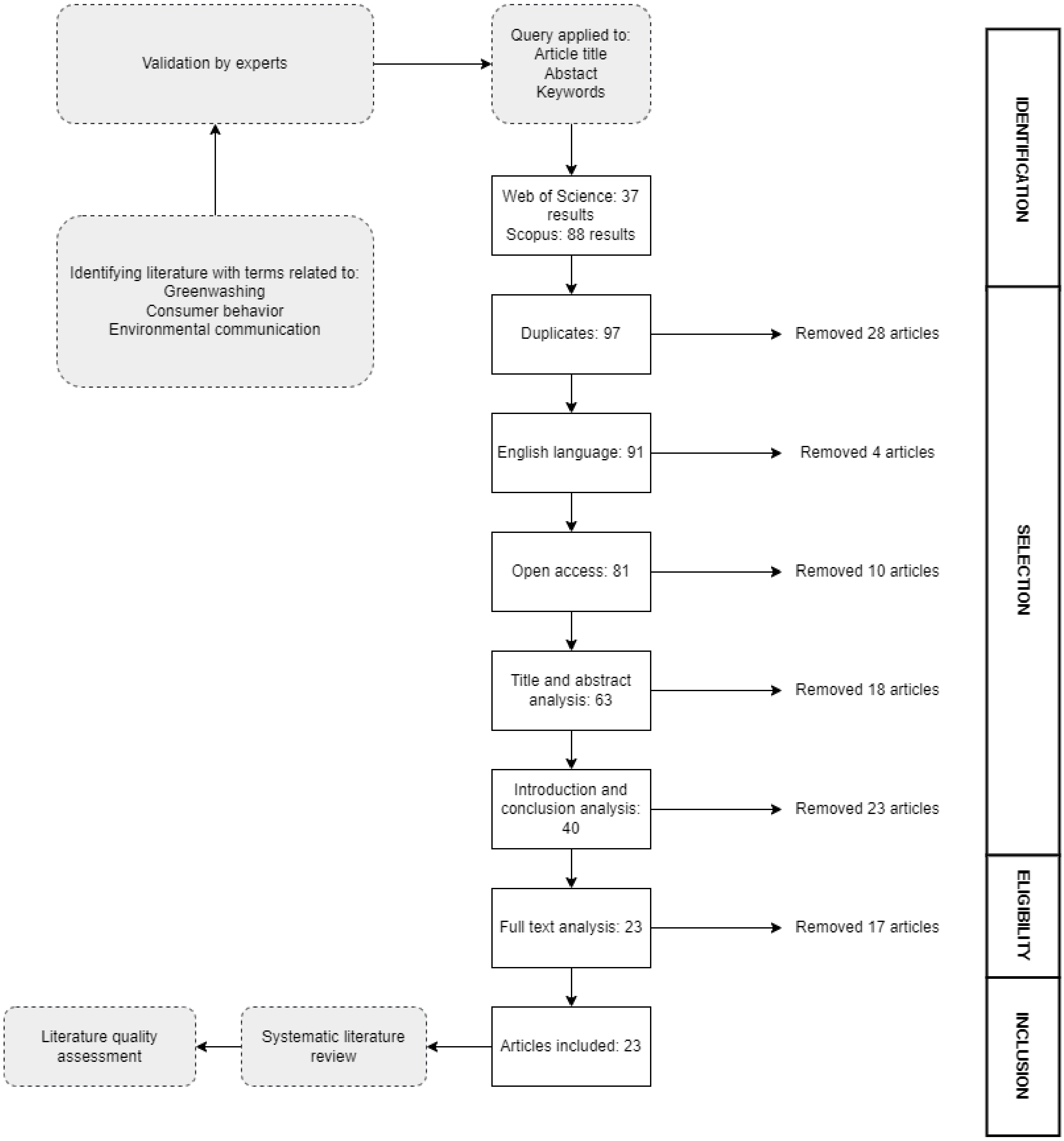

Figure 1 - Article selection process.

2.1.3. Evaluation of articles from the systematic literature review

The final stage of the systematic literature review involves evaluating the literature, where criteria are assigned for each research question. These criteria, summarized in Table 2, determine which articles best address each question. Each article was evaluated based on ten criteria, considering the methodology, variables, extraction techniques, identification and validation of marketing campaigns, factors affecting consumer perception, and marketing strategies presented. This evaluation allowed to determine the quality and significance of the articles, enhancing the clarity and confidence of the systematic literature review results (Kitchenham & Brereton, 2013). The scores range from 0 to 1, with 1 assigned to articles that thoroughly answer the question, 0.5 to articles that partially answer it, and 0 to those that do not.

Table 2 - Criteria for evaluating articles from the systematic literature review.

Questions	Criteria for the evaluation of articles	
i	Q1.1	It defines the methodology used in investigating consumer perception of green marketing campaigns.
	Q1.2	Clearly describe the characteristics of the sample on which the research was based.
ii	Q2.1	Lists explicitly the factors to be considered when investigating consumer perception of green marketing campaigns.
	Q2.2	Explain the impact of each factor when investigating consumer perception of green marketing campaigns.
iii	Q3.1	Describes clearly how the marketing campaigns were identified concerning the practice of greenwashing.
	Q3.2	Lists the sources and forms of validation of campaigns accused of greenwashing.
iv	Q4.1	Describes the factors that can shape consumer perceptions of greenwashing campaigns.
	Q4.2	Clearly describes the influence of each factor.
v	Q5.1	Outlines strategies to reduce consumer perception of greenwashing.
	Q5.2	Distinguishes between marketing strategies for the most polluting industries and others.

2.2. Summary of the articles' contents

Table 3 - Articles included in the systematic literature review.

ID	Year	Title	Journal	Authors	Quotes
1	2021	How Green Trust, Consumer Brand Engagement and Green Word-of-Mouth Mediate Purchasing Intentions	*Sustainability-basel*	Guerreiro, J; Pacheco, M	52
2	2021	Perceived Greenwashing: The Effects of Green Marketing on Environmental and Product Perceptions	*J bus ethics*	Szabo, S; Webster, J	100
3	2021	Communicating Environmental CSR towards Consumers: The Impact of Message Content, Message Style and	*Sustainability-basel*	Christis, J; Wang, YJ	54

ID	Year	Title	Journal	Authors	Quotes
		Praise Tactics			
4	2021	Green but ignored? The irrelevance of television advertisements on energy sustainability in Spain and its impact on consumer perceptions	*Energy res soc sci*	Banares, AB; Silva, MFS; Rodriguez, SR	46
5	2021	Green Talk or Green Walk: Chinese Consumer Positive Word-of-Mouth to Corporate Environmental Actions in Polluting Industries	*Sustainability-basel*	Zhang, JJ; Sun, J	79
6	2019	Is Current Way of Promoting Sustainability, Sustainable?	*J nonprofit public s*	Shabbir Husain, RV; Varshney, S	97
7	2022	Morally transgressive companies and sustainable guidelines: seeking redemption or abusing trust?	*Rausp manag. J.*	Munaier C.G.-E.-S., Miyazaki F.R., Mazzon J.A.	0
8	2022	The transparency paradox: When transparency cues helps or backfires for brands?	*J. Clean. Prod.*	Reck R., Castagna A.C., Shuqair S., Costa Pinto D.	0
9	2022	Greenwashed word of mouth (GWWOM): a vibrant influence on customer green behaviour	*J. Glob. Responsib.*	Singh N., Gupta K., Kapur B.	0
10	2022	How Does Young Consumers' Greenwashing Perception Impact Their Green Purchase Intention in the Fast Fashion Industry? An Analysis from the Perspective of Perceived Risk Theory	*Sustainability*	Lu X., Sheng T., Zhou X., Shen C., Fang B.	0
11	2022	Does Greenwashing Influence the Green Product Experience in Emerging Hospitality Markets Post-COVID-19?	*Sustainability*	Zhang H., Ul Ainn Q., Bashir I., Ul Haq J., Bonn M.A.	0
12	2022	Impact of Greenwashing Perception on Consumers' Green Purchasing Intentions: A Moderated Mediation Model	*Sustainability*	Sun Y., Shi B.	0
13	2022	Are companies using Twitter to greenwash and hide bad environmental performance?	*Energy, ecol. Environ.*	Johnson T.F., Greenwell M.P.	1
14	2022	Greenwashing and Bluewashing in Black Friday-Related Sustainable Fashion Marketing on Instagram	*Sustainability*	Sailer A., Wilfing H., Straus E.	3
15	2022	Determinants of corporate sustainability message sharing on social media: A configuration approach	*Bus. Strategy environ.*	Knight H., Haddoud M.Y., Megicks P.	5
16	2022	Comparing the effects of greenwashing claims in environmental airline advertising: perceived	*Int. J. Advert.*	Neureiter A., Matthes J.	1

ID	Year	Title	Journal	Authors	Quotes
		greenwashing, brand evaluation, and flight shame			
17	2021	Greenwashing in environmental marketing strategy in the brazilian furniture market	*Rev. Econ. Sociol. Rural*	de Alencar Caldas M.V., Veiga-Neto A.R., de Almeida Guimarães L.G., de Castro A.B.C., Pereira G.R.B.	2
18	2020	When consumers learn to spot deception in advertising: testing a literacy intervention to combat greenwashing	*Int. J. Advert.*	Fernandes J., Segev S., Leopold J.K.	12
19	2020	Different Shades of Greenwashing: Consumers' Reactions to Environmental Lies, Half-Lies, and Organizations Taking Credit for Following Legal Obligations	*J. Bus. Tech. Commun.*	de Jong M.D.T., Huluba G., Beldad A.D.	28
20	2018	Consumer Perceptions of Green Marketing Claims: An Examination of the Relationships with Type of Claim and Corporate Credibility	*Serv. Mark. Q.*	Musgrove C.F., Choi P., Chris Cox K.	10
21	2018	The influence of greenwashing perception on green purchasing intentions: The mediating role of green word-of-mouth and moderating role of green concern	*J. Clean. Prod.*	Zhang L., Li D., Cao C., Huang S.	145
22	2018	Misleading Consumers with Green Advertising? An Affect–Reason–Involvement Account of Greenwashing Effects in Environmental Advertising	*J. Advert.*	Schmuck D., Matthes J., Naderer B.	114
23	2018	Making Green Stuff? Effects of Corporate Greenwashing on Consumers	*J. Bus. Tech. Commun.*	De Jong M.D.T., Harkink K.M., Barth S.	48

Table 4 - Summary of the articles included in the literature review.

ID	Summary
1	Explores the impact of perceived greenwashing on green purchase intentions by examining the mediating role of green trust, brand engagement, and green word-of-mouth. The goal is to highlight how important it is for companies to increase green trust and promote green word-of-mouth to increase green purchase intentions.
2	Investigates consumer perceptions of greenwashing by organizations and the impact of such marketing on consumer behavior and product/company perceptions. The study examines the effects of website interactivity and design on perceived greenwashing and its effect on green risk assessment, green value, brand attitudes, purchase intentions, and the impact on consumer happiness.
3	Examines the influence of environmental social responsibility communication on consumer trust, purchase intention, and consumer advocacy. The impact of message content, message style, and complimentary tactics is assessed through an online experiment.
4	Analyzes the use of green marketing strategies by energy companies in Spain and how consumers perceive them. It also discusses the media's role, price and sustainability influence on consumer decisions, and how consumers evaluate companies.
5	Examines the impact of corporate environmental actions, both real and perceived, on consumers' positive word-of-mouth, as well as their cognitive and emotional responses to greenwashing.
6	Examines the underlying components that create an impression of greenwashing and the subsequent impact on attitude toward a green brand. The authors developed a model that explains the formation of attitudes and conducted an empirical analysis to test it.
7	Analyzes the impact of institutional legitimacy on consumers' attitudes of trust and purchase intention toward brands that have previously been involved in moral transgressions and the effect of greenwashing on these attitudes. The results indicate that institutional legitimacy can help regain trust, but perceived greenwashing decreases trust and purchase intention.
8	Examines how brand strength affects consumers' reactions to transparency in corporate social responsibility practices in the fast fashion industry. The results suggest that for well-known fast fashion brands, transparency in CSR messaging may be seen as greenwashing and may decrease trust and that aligning transparency communication with a new brand image is essential.
9	Investigates the impact of green skepticism and green WOM (GWWOM) on green consumer behavior and finds that GWWOM moderates the relationship between green mistrust and consumer green behavior. Suggests further research is needed to understand the strengths of these associations.
10	Examines the effect of perceived greenwashing on green purchase intention in China's fast fashion industry and finds that perceived risk and impulse buying mediate and mediate, respectively.
11	Investigates the influence of green product awareness, greenwashing, and green consumer confusion on green product/service experience, green purchase intention, and WOM intention in the hotel industry in a developing country after COVID-19. The moderating effect of perceived risk is considered.
12	Explores the relationship between consumers' perceptions of greenwashing and their green purchase intentions in China and finds that a perceived sense of betrayal and environmental responsibility play a role. A moderated mediation model is proposed and tested.

ID	Summary
13	Examines the evolution of climate leadership and environmental messaging on social media by UK companies and assesses whether such messaging is a form of greenwashing. While climate leadership and messaging have increased, there is a need for better metrics to measure companies' environmental impact.
14	Investigates the use of greenwashing and bluewashing strategies by sustainable brands on social media during Black Friday and assess the impact on consumers and brands. Suggests small and medium-sized sustainable businesses avoid these practices to maintain credibility.
15	Analyzes the relationship between characteristics of companies' sustainability messages and social media users' intention to share. Identifies conditions influencing message-sharing behavior and offers insights into information quality and source credibility.
16	Examines the impact of greenwashing claims in airline advertisements on brand evaluation, feelings of shame in flying, and the role of consumers' environmental knowledge.
17	Analyzes consumers' perceptions of greenwashing and environmental marketing practices in the furniture industry in Northeastern Brazil, identifying variables that explain the practice of greenwashing and its managerial implications.
18	Studies the effectiveness of a literary intervention to train consumers to evaluate green advertising critically. It confirms consumers' inability to assess truthfulness in green advertising and that a literary intervention can equip them with the necessary knowledge. It focuses on adult consumers and extends the literature on advertising literacy by examining the specific area of ecological advertising.
19	Examines the consequences of greenwashing when consumers discover it. Distinguishes three levels of behavior-based greenwashing and a motivations-based condition. Results suggest that people react similarly to more ambiguous and clear cases of greenwashing.
20	Evaluates the impact of different types of green marketing claims and company trust/friendliness on consumer perceptions of those claims. Results show that substantive claims generate more interest and positive attitudes, while company trust/friendliness plays a moderating role.
21	Investigates the effects of perceived greenwashing on green purchasing intentions, considering the mediating role of green WOM and the moderating role of consumers' green concern. Results suggest that perceived greenwashing has a negative effect on green purchasing intentions and that green concern reinforces this relationship.
22	Investigates greenwashing in advertising and the mediating role of both rational and affective persuasion mechanisms on consumer evaluations of ads and brands. Results show that false claims are perceived as misleading, while vague claims and evocative nature images increase consumer evaluations.
23	Describes experimental research on the impact of greenwashing on consumers. The results indicate that greenwashing has a mixed effect on consumers' attitudes and does not increase their purchase intention. The authors suggest that cognitive dissonance theory could help understand the impact of greenwashing and call for further research to explore the theoretical issues.

2.3. Evaluation of articles from the systematic literature review

In the systematic literature review, the final step involves evaluating the literature based on a set of criteria. These criteria are summarized in Table 2 and are used to determine which articles are most relevant to each research question. Each article is evaluated based on nine criteria, considering factors such as the study's context, variables, methodology, limitations, and contributions. The scoring system ranges from 0 to 1, with a score of 1 given to articles that fully address the research question, 0.5 for partially addressing it, and 0 for not addressing it. This evaluation process helps identify each article's quality and importance and contributes to the reliability of the systematic literature review results (Kitchenham, 2004).

Table 5 - Evaluation of articles from the systematic literature review

ID	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Total
1	1	0.5	1	1	0.5	0	1	1	1	0	7
2	1	0.5	1	1	0	0	1	1	0	0	5.5
3	1	1	1	1	1	0	1	0.5	1	0	7.5
4	1	0	1	0.5	1	0.5	1	0	0.5	0	5.5
5	1	1	1	1	1	0.5	1	1	1	0.5	9
6	1	1	1	1	0.5	0	1	1	1	0	7.5
7	1	1	1	0.5	1	1	0.5	0.5	1	0	7.5
8	1	1	1	1	1	1	1	1	1	0	9
9	1	1	1	1	0	0	1	1	0	0	6
10	1	1	1	1	0	0	1	1	1	0	7
11	1	1	1	1	1	0.5	1	1	1	0	8.5
12	1	0.5	1	1	1	0	1	1	1	0	7.5
13	0.5	1	0.5	0	1	1	0	0	0	0	4
14	1	1	1	1	1	1	1	1	1	0	9
15	1	1	1	1	1	0	1	1	1	0	8
16	1	1	1	1	1	1	1	1	1	0	9
17	1	0	1	0.5	0	0	1	1	0.5	0	5
18	1	0.5	1	0.5	1	0	0.5	0.5	1	0	6
19	1	1	1	0.5	1	1	1	0.5	1	0	8
20	1	1	1	1	1	0	1	1	1	0	8
21	1	1	1	1	1	0	1	1	1	0	8
22	1	1	1	1	1	0	1	1	1	0	8
23	1	1	1	1	1	0	1	1	0.5	0	7.5

2.4. Greenwashing

Greenwashing has become an increasingly prevalent issue in the business world as more and more companies try to capitalize on the growing consumer demand for environmentally friendly products and practices. The term "greenwashing" was first coined in 1986 by Jay Westervelt, who published an essay on hospitality industry practices promoting towel reuse. Since then, greenwashing has been defined in various ways by different authors and researchers, reflecting its multifaceted nature. According to Baum (2012) and Delmas & Burbano (2012), greenwashing refers to misleading consumers regarding a company's environmental practices or performance. Walker & Wan (2012) presented the first economic analysis of greenwashing, defining the phenomenon as "selective disclosure of positive information about a company's environmental or social performance, without full disclosure of negative information on these dimensions, to create an overly positive corporate image." Some authors associate greenwashing with a decoupling behavior, whereby companies engage in symbolic actions that deflect attention or create "green talk" rather than substantive actions to improve their environmental performance.

Companies that engage in greenwashing may create a gap between their symbolic and substantive corporate social actions (Walker & Wan, 2012). In such cases, companies with negative CSR performance may apply positive communication about their CSR performance, thereby misleading consumers. Greenwashing can occur at both the organizational level, where companies mislead consumers regarding their environmental practices, and at the offering level, where companies use misleading environmental claims to refer to the ecological benefits of their offerings. Most research has focused on offering-level claim greenwashing, where companies use false claims, omission of important information, or vague and ambiguous terms to mislead consumers.

In addition to claim of greenwashing (uses textual arguments that explicitly or implicitly refer to the ecological benefits of a product or service to create a misleading environmental claim), (Parguel et al., 2015) introduced the concept of executional greenwashing, which uses nature-evoking elements in advertising to suggest ecological friendliness. Such elements include natural landscapes, endangered animals, and renewable energy sources. These elements may induce false perceptions of a brand's greenness by subtly activating implicit references to nature through imagery, as noted by Hartmann & Apaolaza-Ibáñez (2012).

Overall, greenwashing is a complex and pervasive phenomenon discussed by scholars from various fields. As consumers become increasingly concerned about the environment and seek products and services that align with their values, companies must engage in genuine and substantive efforts to improve their environmental performance rather than simply using greenwashing as a marketing tactic.

2.5. Green marketing

2.5.1. Relevance of green marketing

Green marketing has become a popular tool for companies to appeal to consumers increasingly concerned about the environment. Spanish energy companies have evolved their advertising campaigns to include green elements related to sustainability and environmental protection (Beriain Bañares et al., 2021). However, few direct references were made to green, clean, and renewable energy in energy advertisements (Beriain Bañares et al., 2021). When choosing an energy company, consumers prioritize price over the company's respect for the environment and climate change (Beriain Bañares et al., 2021). This suggests that although consumers have evolved positively regarding pro-environmental knowledge, awareness, attitudes, and behaviors, not every consumer is willing to pay extra for environmentally friendly offerings (ShabbirHusain & Varshney, 2019).

Managers believe that green advertising does not influence consumers purchasing decisions, and the lack of consumer knowledge about the green cause maximizes their lack of engagement (Szabo & Webster, 2021). Green marketing is viewed from different perspectives, including intentional and unintentional greenwashing. Companies need to communicate their green efforts and motivations clearly and honestly, and green concern and responsibility should be embedded in their core values and identity to strengthen green trust (Guerreiro & Pacheco, 2021). This is important because consumers are becoming more aware of the impact of their consumption modes on the environment and are more likely to purchase sustainable and greener products (Zhang & Sun, 2021).

Environmentally conscious consumers will well receive a well-implemented green brand identity, leading to a positive attitude towards the brand (ShabbirHusain & Varshney, 2019). Green advertising is a tool to inform consumers about a company's environmentally friendly products, create positive attitudes toward green businesses, and stimulate demand for green products (Fernandes et al., 2020). However, companies must refrain from misleading green advertising and environmental messages inconsistent with their core values and identity, as this can damage consumer perception and erode green trust (Guerreiro & Pacheco, 2021).

Green marketing is essential for companies to fulfill their social responsibility and appeal to environmentally conscious consumers. However, companies must communicate their green efforts clearly and honestly and ensure that green concern and commitment are embedded in their core values and identity to strengthen green trust. Companies that engage in greenwashing or fail to prioritize the environment in their advertising campaigns risk damaging their reputation and losing consumers who prioritize sustainability and environmental protection.

2.5.2. CSR (Corporate Social Responsibility)

CSR has become increasingly important for companies in recent years. It is an entrepreneurial action beyond mere compliance with legislation and seeks to make positive social and environmental impacts (de Alencar Caldas et al., 2021). With growing concerns about global warming, pollution, deforestation, species extinction, and resource depletion, companies are adopting "green" practices to address these issues (de Jong et al., 2020). However, some companies engage in greenwashing, which involves exaggerating their environmental efforts or making false claims about being environmentally responsible (Szabo & Webster, 2021).

Many companies use third-party certifications and transparent communication to overcome consumer mistrust of green marketing to establish positive consumer relationships (Reck et al., 2022; Szabo & Webster, 2021). Transparency helps to build customer trust and brand attitude, but it can backfire if it is not perceived as authentic (Reck et al., 2022). It is essential for companies to match their transparency communication with their actions and to develop a communication plan to present actual transparency actions and changes (Reck et al., 2022). This is particularly important for well-known brands, as consumers' responses to CSR transparency cues vary depending on brand strength (Reck et al., 2022).

CSR initiatives can positively affect corporate reputation, purchase intentions, and consumer loyalty. In addition, CSR can serve as a buffer during times of crisis and is increasingly expected by stakeholders (de Jong et al., 2020). CSR is a multi-faceted construct that varies among stakeholders, including consumers, employees, suppliers, local communities, and the natural environment (Zhang & Sun, 2021). Two archetypal CSR domains are social and environmental, with the social domain focusing on responsible and sustainable relationships between companies, consumers, employees, and local communities, and the environmental domain emphasizing the link between companies and the natural environment (Zhang & Sun, 2021).

Many organizations have provided suggestions for interactive green components, such as presenting narratives, conveying sustainability visuals, displaying educational materials, and partnering with other green organizations (Szabo & Webster, 2021). Images and videos are the most effective strategies for illustrating a storyline, and infographics are the most successful for efficiently conveying statistics (Szabo & Webster, 2021). Companies can also use blogging and podcasts to provide deeper insights into their campaigns and initiatives and discuss topics relevant to their company culture (Szabo & Webster, 2021). Maintaining a solid social media presence, email marketing, and print were also considered valuable tools for outreach (Szabo & Webster, 2021).

In summary, CSR is increasingly important for companies to address social and environmental issues. While CSR initiatives can positively affect corporate reputation and consumer loyalty, companies must match transparency communication with their actions and avoid greenwashing perceptions. Third-party certifications, transparent communication, and interactive green components can help establish positive consumer relationships. Companies should also use various tools such as social media, blogging, and podcasts to provide deeper insights into their campaigns and initiatives.

2.5.3. Greenwashing perception

Greenwashing is about the information companies provide and the psychological perception it creates in consumers. When a company deliberately releases distorted and misleading information about its environmental performance, it can create cognitive dissonance in consumers' minds. The result is a sense of uncertainty and doubt about the company's true intentions, leading to skepticism about its green claims (Zhang & Sun, 2021).

Consumers experience a perception of greenwashing when they are exposed to vague or false environmental claims made by a company. This feeling can result from the company's marketing tactics, such as using green imagery and buzzwords that may not have substantive backing (ShabbirHusain & Varshney, 2019).

Another way greenwashing perception can occur is through a mismatch between a company's environmental communication and its actual environmental practices. This dissonance can cause consumers to view a company's claims with skepticism and disbelief, leading to a perception of greenwashing.

Ultimately, the perception of greenwashing arises from the consumer's skepticism about a company's environmental communication or performance. Companies must be transparent and authentic in their environmental claims to avoid falling prey to this phenomenon. By being accountable and providing evidence for their claims, companies can build trust with consumers and make a genuine impact in the fight for environmental sustainability (ShabbirHusain & Varshney, 2019).

2.5.4. Claims and communication

Various articles have discussed different types of communication and claims used in green advertising. Companies sometimes make incongruent claims, where their environmental CSR claims do not align with their actual environmental CSR actions (Christis & Wang, 2021). On the other hand, congruent communication occurs when a company accurately describes the action it is taking

(Christis & Wang, 2021). Green advertising can be categorized into different styles, including uniform (CSR claims in line with company actions), apathetic (not involved in CSR talk or action), greenhushing (claiming less action than the company is involved in), and greenwashing (claiming more than the company is doing) (Christis & Wang, 2021). Uniform and apathetic styles are congruent, while greenhushing and greenwashing styles are incongruent (Christis & Wang, 2021).

Some Spanish energy companies have opted for vague claims instead of explicit commitments to avoid accusations of greenwashing (Beriain Bañares et al., 2021). In airline advertising, vague and false claims are common, such as promoting flying as environmentally friendly, making false claims of low carbon emissions or complete elimination of emissions, or omitting crucial information needed for consumers to make responsible consumption decisions (Neureiter & Matthes, 2022). However, compensation claims propose environmental measures and give airlines a reason to claim to be environmentally friendly (Neureiter & Matthes, 2022). There are two types of compensation claims: abstract and concrete (Neureiter & Matthes, 2022). Abstract compensation claims propose future environmental compensation, while concrete compensation claims promote immediate compensation for the environmental impact of flying, which can be directly experienced by consumers (Neureiter & Matthes, 2022). Greenwashing can take many forms, such as making vague and ambiguous claims, omitting important information, including false or fabricated information, or using nature-evoking images to imply environmental benefits (Fernandes et al., 2020). A U.S. magazine content analysis revealed that 43% of ads featured vague or ambiguous claims, and 8% were classified as omissions (Fernandes et al., 2020). Despite this, a recent replication of the study found that more claims were deemed acceptable than deceptive, but misleading claims in advertising are still common.

Furthermore, greenwashing can be divided into behavioral claim and motive (de Jong et al., 2020). Behavioral claim greenwashing refers to the discrepancy between environmental claims and behavior (de Jong et al., 2020). This type of greenwashing is further differentiated between companies that tell the truth, half-lies, and lies (de Jong et al., 2020). On the other hand, motive greenwashing refers to the discrepancy between communicated and real motives for environmentally friendly behavior (de Jong et al., 2020). Motive greenwashing is also differentiated between companies that acted green on their initiative and those that took credit for complying with legal environmental obligations (de Jong et al., 2020).

Additionally, some articles distinguish between different environmental strategies used by companies, such as vocal green (companies that combine good environmental performance with positive communication about their environmental performance), silent green (companies that do not communicate about their good environmental performance), greenwashing (companies that combine bad environmental performance with positive communication about their environmental performance), and silent brown (companies which have bad environmental performance and no communication about environmental performance) (de Jong et al., 2018).

It is also essential to distinguish between substantive environmental actions, which are concrete, visible actions that companies have done or are doing, and symbolic environmental actions, which are superficial, negligible, and easy-to-be-observed environmental gestures aimed to obtain external validation and social support (Zhang & Sun, 2021).

Consumer perception of green communication can significantly impact their purchasing behavior and trust in brands. Research has shown that concrete steps for taking action may not be more effective than environmental CSR communication in encouraging consumers to buy advertised products (Christis & Wang, 2021).

Participants exposed to consumer praise were likelier to trust the brand, while no significant effect of praise tactics was observed (Christis & Wang, 2021). The study also found that attributed intrinsic corporate motives directly predicted trust level, purchase intention, and consumer advocacy. In contrast, no support was found for the proposed moderation effects predicting the value of extrinsic motives (Christis & Wang, 2021).

When it comes to compensation claims, concrete claims did not result in higher levels of perceived greenwashing compared to the control condition, while abstract, vague, and false claims resulted in significantly higher levels of perceived greenwashing than the control condition (Neureiter & Matthes, 2022). Vague and false claims had a higher level of perceived greenwashing compared to concrete compensation claims but not compared to abstract compensation claims (Neureiter & Matthes, 2022). The study suggests that consumers perceive green airline ads as incongruent with their mental representation of flying being environmentally harmful, leading to increased cognitive attention and perceptions of greenwashing (Neureiter & Matthes, 2022). However, consumers accept green claims offering concrete green compensations, possibly due to a lower incongruence between the ad and existing mental representations (Neureiter & Matthes, 2022).

Another study (de Jong et al., 2020) found that the motive for greenwashing generally did not matter for participants, indicating that it did not matter whether the green initiatives were self-initiated or merely reflected compliance with legal obligations. However, the company's motives' truthfulness was only meaningful if it put its environmental claim into practice.

When it comes to the severity of greenwashing, telling lies and half-lies about environmentally friendly behaviors has detrimental effects on corporate reputation constructs, including environmental performance, perceived product and service quality, and perceived financial performance (de Jong et al., 2020). The negative effects also occur in a less severe and less obvious case of greenwashing (the half-lies condition), and the effects are most potent on environmental performance but also significant and practically meaningful on the perceived quality of products and services and perceived financial performance (de Jong et al., 2020).

The study suggests that people care more about companies' dishonesty about environmental policies than about the extent of misalignment between claims and behaviors, as partial lies can have the same detrimental effects as complete lies (de Jong et al., 2020). The motive for greenwashing only adds to the reputational damage if the company is not guilty of behavioral claim greenwashing (de Jong et al., 2020). If the company was guilty of behavioral-claim greenwashing, the motive of greenwashing did not have a significant effect (de Jong et al., 2020).

Green marketing claims can significantly impact consumer perception, with substantive claims generating higher interest levels and positive attitudes toward the company than posturing claims (Musgrove et al., 2018). High levels of trustworthiness and likability can also improve consumer perception, but they may fail to reduce skepticism in some cases (Musgrove et al., 2018). However, even for credibility-challenged companies, substantive green marketing claims can produce a "green halo" effect, suggesting that they can be especially powerful in generating positive consumer response (Musgrove et al., 2018).

While false claims significantly increase perceived greenwashing, vague claims do not necessarily have the same effect (Schmuck et al., 2018). Results suggest that vague claims in green advertising are unrelated to perceived greenwashing and do not increase consumers skepticism compared to non-deceptive claims (Schmuck et al., 2018). False claims are perceived as more deceptive and factually wrong than vague claims, indicating that they substantially impact consumer perception (Schmuck et al., 2018).

Greenwashing can also significantly impact how consumers view companies' environmental claims and performance (de Jong et al., 2018). Companies that explicitly communicate interest in environmental issues can create a more favorable image than those that neglect the environment. However, consumers appreciate companies that communicate positive environmental behavior, those that are silent about it, and those that neglect it equally (de Jong et al., 2018). Greenwashing companies receive lower scores on perceived integrity and environmental performance than vocal green and silent green companies but more positive scores than silent brown companies (de Jong et al., 2018).

These findings suggest that companies should be cautious in their green marketing claims and prioritize substantive claims that accurately reflect their environmental performance rather than making vague or false claims that could lead to consumer skepticism and perceptions of greenwashing. Companies can build trust and positive consumer perceptions by demonstrating a commitment to environmental responsibility and communicating it honestly and transparently.

Consumer perception plays a crucial role in shaping the success or failure of green communication efforts by companies. According to (Zhang & Sun, 2021), consumers are more likely to respond positively to substantive environmental actions taken by companies rather than symbolic actions that only portray a green image without making any actual environmental improvements. Implementing strict environmental guidelines and management can help companies build credibility and gain consumer trust. Vague or deceptive advertising, on the other hand, can lead to negative associations and harm the brand's reputation (ShabbirHusain & Varshney, 2019).

The credibility of a company's message in green advertising is critical to consumer perception. Consumers take the ad message as a source of information and are likelier to believe and trust it if perceived as truthful and believable (ShabbirHusain & Varshney, 2019). Green ad skepticism has a significant negative impact on green brand attitudes, and false environmental claims by companies can increase consumers' perceived greenwashing, which harms their attitudes toward ads and brands (Fernandes et al., 2020; ShabbirHusain & Varshney, 2019). Therefore, companies must disclose all necessary information to support their green product's environmental benefits and performance claims through various communication platforms (ShabbirHusain & Varshney, 2019).

The study by (Fernandes et al., 2020) highlights that consumers' inability to distinguish between what the FTC (Federal Trade Commission) defines as deceptive vs. acceptable environmental claims emphasizes the need for educating consumers on how to read and decode green messages. Difficulty in verifying advertising claims can lead to skepticism, distrust, or disbelief in advertising (Fernandes et al., 2020). Consumers cannot evaluate the extent of truthfulness or deception in green advertising (Fernandes et al., 2020). Therefore, companies should make their green communication efforts as transparent and accurate as possible to build consumer trust and improve brand perception.

In summary, companies that take substantive environmental actions and communicate them transparently to their consumers are more likely to gain their trust and build a positive brand image. Vague or deceptive advertising can harm the brand's reputation, while greater message credibility can positively impact green brand attitude. However, consumers' inability to distinguish between acceptable and deceptive claims highlights the need for educating them on how to decode green messages and verify advertising claims.

In recent years, companies have increasingly adopted Corporate Social Responsibility (CSR) practices to communicate their environmental commitment. However, the challenge for companies lies in effectively communicating these initiatives to consumers while maintaining credibility and avoiding skepticism (Christis & Wang, 2021). Research has shown that consumers are more likely to respond positively to substantive environmental actions rather than symbolic actions that merely aim to show a green image (Zhang & Sun, 2021). This highlights the importance of synchronizing communication with actual environmental CSR actions and adopting a uniform message style to create trust and enable consumer advocacy (Christis & Wang, 2021).

One practical approach to achieving this is focusing on action-message congruency, transparency, and intrinsic or extrinsic motives rather than emphasizing praise tactics or specific CSR content (Christis & Wang, 2021).

Companies must refrain from using greenwashing techniques as incongruence between actions and messages can lead to lower trust levels and decreased consumer advocacy (Christis & Wang, 2021). The key is to have a truthful and transparent approach to environmental responsibilities, which can protect companies from criticism and establish long-term bonds with consumers (Christis & Wang, 2021).

Effective communication is essential to create consumer affiliation with a product or company, leading to consumer advocacy (Christis & Wang, 2021). CSR should not be treated as a unidimensional concept but also include consumer advocacy (Christis & Wang, 2021). Therefore, companies must adopt a uniform communication style that aligns their actions and messages (Christis & Wang, 2021). In addition, corporate sustainability missions can be better achieved through transparency, message congruency, and intrinsic aspirations for the involved environmental CSR cause (Christis & Wang, 2021).

To avoid greenwashing, companies should ensure that their behavior aligns with environmental communication and be honest about their motives (de Jong et al., 2020). Consumers are disadvantaged in identifying deceptive environmental claims (Fernandes et al., 2020). Thus, providing consumers with expertise in identifying greenwashing can help them evaluate green messages and decrease companies' motivation to engage in greenwashing (Fernandes et al., 2020). In conclusion, companies should maintain transparency, communicate truthfully, and align actions and messages to establish long-term trust and advocacy with consumers in their environmental CSR initiatives.

2.5.5. Greenwashing and emotions

Greenwashing, using misleading marketing tactics to promote environmentally friendly products or services, can negatively impact consumer emotions (Szabo & Webster, 2021). Perceived greenwashing has been linked to decreased happiness, as measured through facial expressions (Szabo & Webster, 2021). Understanding the psychological mechanisms underlying consumer attitudes and behaviors toward corporate environmental engagements is essential for the success of companies (Zhang & Sun, 2021). Cognitive analysis is necessary to identify greenwashing, particularly in industries that have traditionally faced criticism for polluting the environment (Zhang & Sun, 2021).

Moral emotions, such as other-condemning emotions, are critical in understanding the psychological mechanism behind consumer positive word-of-mouth toward corporate environmental actions (Zhang & Sun, 2021). Other condemning emotions, such as contempt, anger, and disgust, are generated when criticizing the characteristics or behavior of others (Zhang & Sun, 2021). These emotions decrease consumer-positive word-of-mouth and mediate the relationship between corporate environmental actions and consumer-positive word-of-mouth (Zhang & Sun, 2021).

The study suggests that moral emotions are more automatic and generic than cognitive reactions and can trigger moral judgment of specific events, influencing consumer responses (Zhang & Sun, 2021). Practitioners should avoid evoking consumer disgust and contempt when delivering corporate symbolic environmental actions (Zhang & Sun, 2021).

Overall, understanding the psychological mechanisms underlying consumer responses to corporate environmental actions is crucial for companies. Perceived greenwashing can lead to negative emotions, while genuine environmental actions can generate positive consumer attitudes and behaviors. Practitioners should avoid greenwashing and adopt authentic, substantive environmental actions that align with consumer values and generate positive emotions. By doing so, companies can establish strong relationships with consumers and achieve long-term success.

2.6. Greenwashing and social media

The use of social media and its impact on corporate sustainability communications has been the focus of multiple studies. (Lyon & Montgomery, 2013) state that traditional media plays a vital role as it can serve as a watchdog over companies, but its ability to challenge the legitimacy of companies has been limited due to control by elites. On the other hand, social media provides more accessible and free platforms for consumers and activists to voice their opinions and allows for symmetrical two-way communication between companies and consumers. The impact of social media on corporate greenwashing varies based on a company's environmental reputation and the degree of greenness. However, the authors theorize that the advent of social media may reduce corporate greenwashing. Companies with good green reputations should reduce promotion when they have bad news, and companies in clean industries or those that focus on green products should communicate via social media.

A study by (Johnson & Greenwell, 2022) of hundreds of UK companies explored the relationship between environmental performance, environmental messaging, and financial performance and found that companies with higher environmental messaging and better environmental performance may be more profitable.

A study by (Sailer et al., 2022) of sustainable fashion marketing on Instagram aimed to identify greenwashing and bluewashing strategies in Black Friday-related content and the most relevant predictors of brand and sustainability evaluation by consumers found that companies commonly used CRM (offering to make a small donation to a charity for each purchase) strategies, rebranding, offering sustainability-related promotional gifts, and using sustainability-related imagery to appeal to consumers.

The research by Knight et al. (2022) on sustainable business strategy, marketing, and consumer behavior aimed to understand the key determinants of social media message sharing in the context of corporate sustainability communications. The study adopted a configuration approach to understanding the factors that determine social media users' willingness to share corporate sustainability messages and found that high information relevance and high information accuracy appear to be substituted for one another. Both need to be complemented with high source expertise and high trustworthiness to reach viral sharing. The authors found that companies with strong environmental reputations benefit from using social media to communicate their sustainability efforts and that credibility cues in the form of certification strengthen consumers' beliefs. The quality of the argument in the message is more critical if source credibility is not fully established. However, the presence of both source expertise and trustworthiness overcomes the requirement for all three information quality dimensions. The study supports previous findings that the message must be meaningful and relevant for consumers to participate in viral communication.

Word-of-mouth (WOM) has become increasingly important in shaping consumers purchasing behavior and opinions, particularly in the age of the internet and social media (Singh et al., 2022). Consumers are likelier to spread positive WOM about substantive action than symbolic corporate environmental actions (Zhang & Sun, 2021).

Green skepticism is a factor that can influence greenwashing word-of-mouth (GWWOM) and ultimately lead to a shift in consumer behavior (Singh et al., 2022). This highlights the importance of companies addressing any negative WOM that may harm their brand image and collecting suggestions for green improvements to maintain customer satisfaction and meet environmental expectations (Guerreiro & Pacheco, 2021).

Regarding the hospitality industry, positive experiences with green lodging products can result in positive repurchase intentions and WOM (Zhang et al., 2022). However, perceived risk about green lodging products can moderate consumer perceptions and influence their intent to repurchase and provide positive WOM (Zhang et al., 2022). Social media campaigns and eco-centric integrated marketing communications can strategically educate consumers about green products and services and influence their purchasing behavior and WOM (Zhang et al., 2022).

Greenwashing perceptions can also impact consumer purchasing intentions and WOM (Zhang et al., 2022; Zhang et al., 2018). However, the mediating effect of green WOM suggests that companies should strengthen it to encourage green purchasing intentions and reduce the negative impacts of greenwashing perceptions. Monitoring social media platforms for customer feedback and engaging with consumers can also be an effective strategy for companies to maintain consumer satisfaction and meet environmental expectations (Guerreiro & Pacheco, 2021; Zhang et al., 2022).

The impact of social media on corporate environmental communication and greenwashing reduction will vary depending on a company's environmental reputation, degree of greenness, credibility, and expertise in sustainability issues. WOM is crucial in shaping consumer behavior and opinions towards green products and services, highlighting the importance of companies promoting substantive environmental actions and addressing negative feedback to maintain a positive brand image and encourage green purchasing behavior and positive WOM.

2.7. Greenwashing and NLP

Natural language processing (NLP) is a branch of artificial intelligence concerned with the interaction between computers and human (natural) languages. The combination of greenwashing and NLP is an emerging research area that aims to explore the relationship between environmental messaging and NLP algorithms. Such research seeks to identify the extent to which NLP algorithms can be used to detect and mitigate greenwashing in environmental messaging. However, the body of research in this area is scarce, indicating significant room for further investigation and development.

Nonetheless, one study (Johnson & Greenwell, 2022) aimed to fill this gap by examining UK companies' environmental performance and messaging. The study collected data from the CDP (2020) and from Twitter, where the last 3200 tweets from 287 companies were downloaded and searched for environmental terms such as "sustainability" and "renewable energy." The study aimed to explore the relationship between a company's environmental performance and messaging, as well as its financial performance, and to determine the potential impact of environmental messaging on a company's financial performance.

The researchers manually searched for the Twitter handle of each of the 510 companies and removed uncertain accounts, leaving 287 companies for analysis. A logistic mixed model was used to determine the change in environmental messaging over time and to model the presence of environmental terms in tweets against "years since 2010". The study also characterized the overall word use by identifying the 150 most used words across the Twitter profiles of UK companies and removed stop words and stemmed words to get the results.

The results showed that environmental messaging was rare, with only one word, "sustain" related to the environment found in the 150 most used words across all tweets. Despite this, environmental messaging did increase over time. The growth in environmental messaging was highly variable across sectors, with the fastest growth in electric utilities and the slowest growth in construction. Paper and forestry were the sectors with the most environmental messaging across companies and years. The study found that environmental messaging on Twitter increased over time, particularly in companies with higher climate leadership scores, suggesting that companies are genuinely committed to being environmentally sustainable. However, the sector-specific variation was substantial, meaning what may be adequate environmental messaging in one sector could be over or under-promoting in another. The study also found that the number of followers did not impact environmental messaging but positively impacted climate leadership.

In conclusion, the study found that despite the increase in environmental messaging on Twitter by companies over time, these messages were still rare, and there is room for companies to promote their pro-environmental actions further. Companies with higher climate leadership scores were more likely to engage in environmental messaging, suggesting that companies are promoting genuine efforts to be more environmentally sustainable. The scarcity of environmental messaging suggests that companies may be prioritizing other themes, and the study is not meant to determine what level of messaging is acceptable and proportionate, as it only assesses if companies are promoting their environmental actions proportionately to their climate leadership.

2.8. Answer to descriptive questions

The descriptive questions can now be answered based on the results obtained from the systematic literature review conducted:

i) **What methods were used to investigate how consumers perceive green marketing campaigns that are accused of greenwashing?**

Various methods have been used to investigate consumer perception of greenwashing in ecological marketing campaigns, including surveys (Szabo & Webster, 2021), questionnaires (Beriain Bañares et al., 2021; Christis & Wang, 2021; ShabbirHusain & Varshney, 2019; H. Zhang et al., 2022), controlled experiments (de Jong et al., 2020; Guerreiro & Pacheco, 2021; Neureiter & Matthes, 2022; Sun & Shi, 2022; Szabo & Webster, 2021), and interviews (de Alencar Caldas et al., 2021; Szabo & Webster, 2021). These methods explore how green trust, consumer brand engagement, and green word-of-mouth affect consumers' expectations of greenwashing and their purchasing decisions. The studies typically measure consumer perceptions of deceptive green marketing, environmental risk, environmental value, attitudes toward the brand, purchase intent, and happiness (Szabo & Webster, 2021). They also assess the impact of CSR message content (Christis & Wang, 2021), message style (Christis & Wang, 2021; de Jong et al., 2020; Schmuck et al., 2018), and praise tactics on consumers' trust (Christis & Wang, 2021), purchase intention, and consumer rights (Beriain Bañares et al., 2021; Munaier et al., 2022; Musgrove et al., 2018; Singh et al., 2022). In addition, the investigations analyze consumers' overall perception of companies and their commitment to environmental protection, as well as their attitudes toward green brands, ecological awareness, and green consumption (Guerreiro & Pacheco, 2021; Sailer et al., 2022; ShabbirHusain & Varshney, 2019).

ii) What factors should be considered when assessing consumer perception of green marketing campaigns accused of greenwashing?

Factors to consider include environmental beliefs (Szabo & Webster, 2021), green risk (Lu et al., 2022; Oliveira et al., 2019; Szabo & Webster, 2021), attitudes toward the brand (Musgrove et al., 2018; Szabo & Webster, 2021), purchase intention (Guerreiro & Pacheco, 2021), consumer confidence, consumer advocacy (Christis & Wang, 2021), overall brand perception (Beriain Bañares et al., 2021; Munaier et al., 2022), perceived quality (Beriain Bañares et al., 2021), perceived trust (Beriain Bañares et al., 2021; Christis & Wang, 2021; Guerreiro & Pacheco, 2021; Oliveira et al., 2019; Reck et al., 2022), perceived loyalty, perceived image, perceived price-quality ratio, perceived satisfaction, and perceived preference for the brand's products and services compared to the competition (Beriain Bañares et al., 2021). Other factors to consider include skepticism towards green advertisements, the credibility of the company's green message, the perceived effectiveness of the company on environmental impact, individual consumer characteristics (ShabbirHusain & Varshney, 2019), brand size, information transparency (Reck et al., 2022), green skepticism (Singh et al., 2022), green word-of-mouth communication (Guerreiro & Pacheco, 2021; Singh et al., 2022; L. Zhang et al., 2018), perceived risk, impulse buying, consumer green awareness, consumer green confusion (Lu et al., 2022; Neureiter & Matthes, 2022; Oliveira et al., 2019; Reck et al., 2022), integrity and reliability of the company's environmental practices, fraudulent environmental claims, vagueness or unverifiability of green claims (de Alencar Caldas et al., 2021; Sun & Shi, 2022), the use of appropriate instruments to measure environmental performance and corporate environmental communication, and the quality and credibility of the message (De Jong et al., 2018; Johnson & Greenwell, 2022; Knight et al., 2022; Musgrove et al., 2018).

iii) **How were the green marketing campaigns accused of greenwashing identified/constructed?**

Various methods were used to identify marketing campaigns and brands accused of greenwashing. These methods include analyzing advertising campaigns broadcast on television (Beriain Bañares et al., 2021), creating fictitious stimulus materials (de Jong et al., 2020; Musgrove et al., 2018; Szabo & Webster, 2021; J. Zhang & Sun, 2021), using Twitter and Instagram to collect data (Johnson & Greenwell, 2022; Sailer et al., 2022), searching for content on social media platforms (H. Zhang et al., 2022), and using specific examples and pilot studies (Fernandes et al., 2020; Munaier et al., 2022; Neureiter & Matthes, 2022; ShabbirHusain & Varshney, 2019; L. Zhang et al., 2018). For instance, in one experiment, participants interacted with different versions of an online page with different interactive green components to communicate product information for brands such as Nestle and Apple (Guerreiro & Pacheco, 2021). In contrast, in another study, a fictional tea brand called "HerbaLove" was created for participants to analyze through an "About" Facebook profile, an independent NGO report on the company's environmental social responsibility activities, and a HerbaLove advertisement (Christis & Wang, 2021).

iv) **What conclusions were drawn regarding consumer perception, and what factors influenced perception?**

The conclusions drawn regarding consumer perception are that perceived greenwashing negatively affects green purchase intentions through various indirect pathways, including green trust and green word-of-mouth. Green word-of-mouth and green trust are essential factors in the relationship between perceived greenwashing and green purchase intentions (Guerreiro & Pacheco, 2021). An effective environmental CSR communication strategy can create consumer affiliation with a product or company at a deeper level, such as stimulating consumer advocacy. Companies must adopt a uniform communication style involving synchronizing the company's environmental CSR actions and messages to achieve effective CSR communication rather than adopting greenwashing techniques that provoke negative reactions among consumers (Christis & Wang, 2021). Consumers attitudes towards a green brand are influenced by their skepticism towards green advertisements (ShabbirHusain & Varshney, 2019), the credibility of the company's green message (Sailer et al., 2022), perceived value (de Alencar Caldas et al., 2021), and awareness of the company's environmental impact (de Jong et al., 2020). Using greenwashing and bluewashing strategies can mislead consumers and negatively affect brand evaluation among sustainability-conscious consumers (Neureiter & Matthes, 2022). Consumers prefer environmental products and are aware of greenwashing but do not always act against it. Only honest and transparent communication about environmentally friendly behaviors has positive feedback, and specific and substantial green marketing messages decrease skepticism. Concrete claims are not perceived as greenwashing, while vague and false claims are. Factors influencing perception include literacy interventions, the nature of claims in advertisements or product packaging, company trustworthiness and likeability, green concern, and participants' state of mind (Schmuck et al., 2018).

v) What are the strategic recommendations for running a green marketing campaign?

The strategic recommendations include avoiding greenwashing activities, strengthening green word-of-mouth, incorporating environmental concerns into company values and identity, communicating green efforts clearly and honestly, monitoring social platforms for insights (Guerreiro & Pacheco, 2021), focusing on substantive rather than symbolic environmental actions, avoiding triggering negative emotions in consumers, being transparent about environmental efforts (J. Zhang & Sun, 2021), providing evident proof of environmental claims (ShabbirHusain & Varshney, 2019; J. Zhang & Sun, 2021), increasing consumer awareness of environmental impact (ShabbirHusain & Varshney, 2019; H. Zhang et al., 2022), embracing social agendas in institutional legitimization actions (Munaier et al., 2022), developing a transparency plan to increase congruence between brand image (Reck et al., 2022), products, and communication (Lu et al., 2022), taking substantial sustainability measures (H. Zhang et al., 2022), promoting authentic communication without greenwashing (Sun & Shi, 2022), reducing perceived risk for consumers (Lu et al., 2022), providing certifications for eco-friendly offerings, promoting consumer awareness of sustainability (Knight et al., 2022), using social media to communicate with target audiences, disclosing environmental practices promptly, avoiding bluewashing (Neureiter & Matthes, 2022), planning a more subtle approach to sustainability communication strategies on social media (Musgrove et al., 2018), focusing on truthful and accurate environmental claims (Fernandes et al., 2020), updating industry guidelines for environmental marketing claims, providing clear and accurate information about offset claims (Oliveira et al., 2019), and implementing tools to help consumers detect misleading messages (L. Zhang et al., 2018). Additionally, companies should involve marketing managers in green initiatives from the beginning, be honest and transparent about environmentally friendly practices, follow FTC guidelines for green marketing, communicate green messages effectively to consumers, encourage and educate consumers to increase their environmental concerns, monitor consumers opinions about green products, avoid false green marketing claims, and strengthen green marketing regulations to gain consumers trust and increase purchase intentions (Schmuck et al., 2018).

CHAPTER 3

Methodology

CRISP-DM (Pete et al., 2000) is a widely used framework for developing and implementing data mining projects. The methodology comprises six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The CRISP-DM methodology is designed to be flexible and iterative, allowing for cycles of refinement and adjustment as the project progresses. It provides a systematic approach to data mining projects, which guides practitioners through the various stages of a project, from initial business understanding to final deployment. The framework is widely used in multiple industries, including healthcare, finance, retail, and telecommunications, as it provides a structured approach to data mining projects that helps ensure project success.

3.1. Business Understanding

The business understanding stage is a critical phase of the CRISP-DM framework in a research investigation. It is the first phase of the framework and sets the foundation for the entire project. This stage aims to gain a comprehensive understanding of the research area, identify the problem that needs to be addressed, and define the objectives and scope of the study. In a research investigation, the business understanding stage involves exploring the research area to understand the problem at hand thoroughly. This includes reviewing existing literature and identifying gaps in knowledge. Through this process, the researcher can identify the research question, objectives, and the scope of the investigation. In this dissertation, the business understanding phase corresponds to what is presented in the introduction and literature review chapters. This section provides an overview of the research area, defines the problem being addressed, and outlines the objectives and scope of the study.

3.2. Data Understanding

Researchers gather and explore the available data in the data understanding stage to better understand its structure, content, quality, and potential limitations. This stage involves various activities, such as data collection, integration, exploration, and cleaning. The primary objective of this stage is to gain insights into the data, understand its characteristics, and identify any issues that may affect the validity and reliability of the results. By doing so, researchers can make informed decisions about preprocessing and transforming the data to prepare it for the modeling phase. The data collection for this research investigation was conducted through Eco-bot.net[1]. The platform's system is designed to collect, visualize, and flag corporate greenwashing ads and content across Facebook, Instagram, and Twitter. Once the system identifies corporate tweets that contain greenwashing content and verifies them through its in-house team of journalists, the data is displayed online and flagged on the social media platform where it was found.

A manual data collection process was performed to gather data on corporate greenwashing from Twitter. This involved scraping the corporate tweets from the website and utilizing Twitter's search functionality to obtain their respective tweet IDs. Subsequently, the data collection process was refined by leveraging the Twitter API with an academic research license. All replies to the identified corporate tweet were collected using a Python script. Several steps were taken to ensure data quality and completeness for the research investigation. Firstly, replies from the Eco-bot.net Twitter account were not collected to avoid any biases that might have influenced the data. Secondly, corporate tweets without replies were also not collected as they would not have contributed to the research objectives. These steps made the collected data more focused, relevant, and less prone to biases or inaccuracies.

3.3. Data Preparation

The data preparation stage is the third and critical step in the data mining process. In this stage, the focus is on preparing the data for analysis in subsequent stages. This involves selecting the appropriate data, transforming the data, and creating new variables. The data preparation stage is crucial for ensuring that the data used for analysis is of high quality and appropriate for the modeling techniques used in subsequent stages. The output of the data preparation stage is a prepared dataset ready for modeling in the next stage of the CRISP-DM model.

3.3.1. Variable creation

As part of the data collection process, several variables were created. To understand the sector of the companies involved, the Morningstar global equity classification structure[2] was used to categorize them. This allowed to categorize companies into sectors such as Technology, Financial Services, Energy, etc. In addition, a CDP[3] score was collected for all available companies in 2021. CDP, which stands for Carbon Disclosure Project, is a global not-for-profit organization that collects information from companies and cities on their impacts and dependencies on the world’s natural resources and their strategies for managing these. The organization does this for investors,

---

[1] https://eco-bot.net/
[2] https://indexes.morningstar.com/resources/PDF/Methodology%20Documents/SectorArticle.pdf
[3] https://www.cdp.net/en/

purchasers, and governments who want to understand better how these entities are performing regarding the environment. A CDP score provides a snapshot of a company's disclosure and environmental performance. CDP uses a scoring methodology to incentivize companies to measure and manage environmental impacts through one or more of their climate change, forests, and water security questionnaires. The CDP score is divided into five categories: Disclosure (D-/D score), Awareness (C-/C score), Management (B-/B score), Leadership (A-/A score), and Failure to disclose (F score). The Disclosure level is the starting point for companies that want to demonstrate that they have begun their environmental journey. This level features both D and D- scores, and companies are awarded roughly one point per data point provided. To score a D over a D-, companies need to have disclosed a more extensive set of information. The Awareness score measures the comprehensiveness of a company's evaluation of how environmental issues intersect with its business and how its operations affect people and ecosystems. The level of awareness is measured through a C or C- score, with the differentiator being the level of awareness a company has shown in its response. The Management level indicates that a company is showing some evidence of managing its environmental impact but is not undertaking actions that mark it as a leader in its field. Companies that score a B have addressed the environmental impacts of their business and ensured good environmental management. To earn an A score, companies must show environmental leadership, disclosing action on climate change, deforestation, or water security. They must demonstrate best practices in strategy and action as recognized by frameworks such as the TCFD (Task Force on Climate-related Financial Disclosures) and the Accountability Framework. An F score is given when a requested company fails to disclose through CDP. The CDP score is based solely on activities and positions disclosed in the CDP response. It, therefore, does not consider actions or activities not mentioned in the CDP response.

Similarly to Johnson & Greenwell (2022), the CDP score ordinal scale was also converted into a numeric integer scale ranging from 8 to 0, where A = 8, A- = 7, B = 6, B- = 5, and so forth up to F = 0. It is important to note that CDP scores alone are not a comprehensive metric of a company's level of sustainability or 'green-ness' but instead indicate the level of action reported by the company to assess and manage its environmental impacts during the reporting year.

Corporate tweets were analyzed based on declarations and actions and labeled according to various categories using a binary system. The process drew upon definitions provided by several authors in their respective research.

Corporate tweets were first labeled according to their declarations of abstract compensation claims, concrete compensation claims, vague claims, and false claims using Neureiter & Matthes (2022) definitions. Abstract compensation claims were defined as claims that propose future environmental compensation for the environmental impact of flying without consumers being able to observe the trade-off directly. In contrast, concrete compensation claims were said to promote immediate compensation for the environmental impact of flying. Vague claims, on the other hand, were described as overly ambiguous and too unspecific to make a well-informed conclusion about the green character of the service. False claims, as the name suggests, include outright lies.

Corporate tweets were also labeled according to their actions, using a binary system that distinguished substantive actions from symbolic actions (Zhang & Sun, 2021). Substantive actions were described as concrete and visible actions in management goals, organizational structures, and social institutionalization initiatives, where companies discuss their environmental responsibility in terms of what they are doing now or what they have done. Symbolic actions, on the other hand, represented an array of superficial, negligible, and easy-to-be-observed environmental gestures aiming to obtain external validation and social support, such as establishing an environmental supervision committee or using green labels or trademarks.

Furthermore, corporate and consumer praising in corporate tweets were also distinguished, as defined by Christis & Wang (2021). Corporate praising involves praising the company for its social actions, such as its sustainability initiatives, while consumer praising involves praising consumers for their sustainable behavior or purchases.

Finally, corporate tweets were flagged for mentioning net zero or related terms, using a binary system where "1" denoted inclusion and "0" indicated exclusion. As Fankhauser et al. (2022) point out, net zero has multiple interpretations, but the general concept is the balance between releasing and removing carbon dioxide into the atmosphere. They also state that governance, accountability, and reporting mechanisms are currently inadequate to assess the credibility of net zero pledges. They point out that long-term ambition is often not backed up by sufficient near-term action. They also note that many entities have not yet set out detailed plans to achieve their pledges and are opaque about the role of carbon offsets in place of cutting their emissions. They argue that some of these offsets' environmental and social integrity is questionable. As a result, some advocates have accused these pledges of amounting to little more than ‘greenwashing.’

To ensure accuracy and avoid biases, OpenAI's ChatGPT[4] was used to label corporate tweets according to the various categories and definitions mentioned above. All responses provided by ChatGPT were manually verified, and its responses were accurate and consistent.
The following diagram outlines the data collection process.

[4] https://openai.com/blog/chatgpt (March 14, 2023)

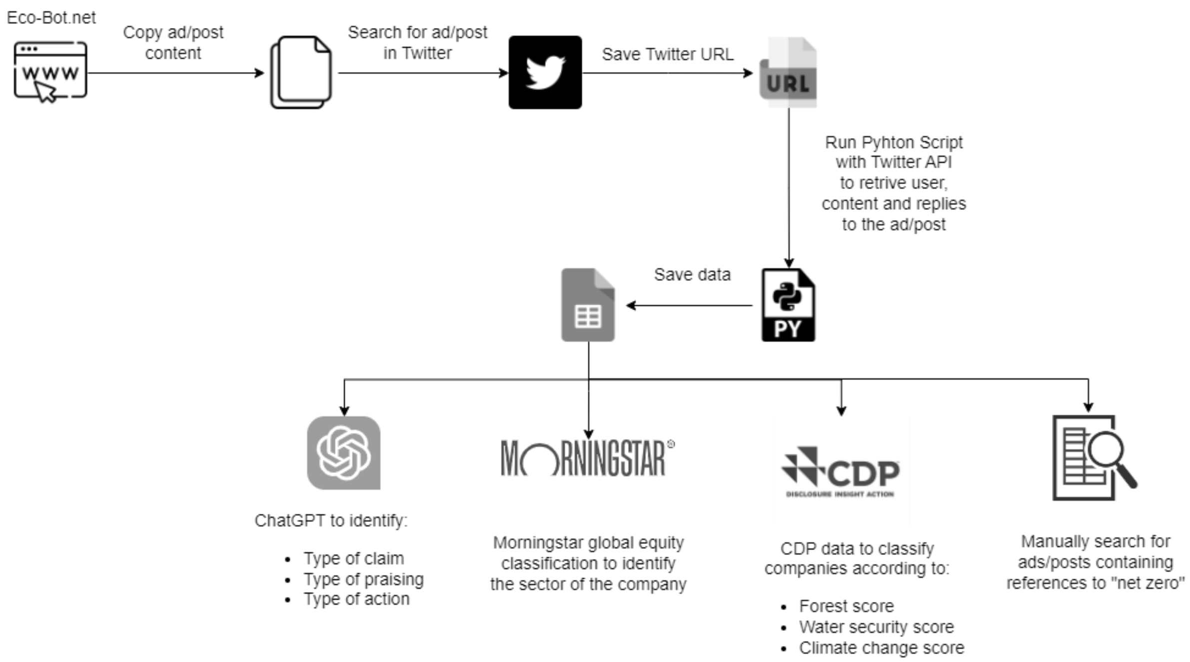

Figure 2 - Data collection process

3.3.2. Descriptive analysis

The data collection process resulted in 451 corporate tweets and 7017 replies to those corporate tweets divided by 131 different companies and ten industry sectors.

The sectors represented in the data set are Basic Materials, Communication Services, Consumer Cyclical, Consumer Defensive, Energy, Financial Services, Health Care, Industrials, NGO and others, and Technology. Basic materials represent companies manufacturing chemicals, building materials, and paper products. This sector also includes companies engaged in commodities exploration and processing. Communication Services represent companies that provide communication services using fixed-line networks or wireless access and services. This sector also includes companies providing internet services such as access, navigation, and internet-related software. Consumer Cyclical includes retail stores, auto & auto parts manufacturers, companies engaged in residential construction, lodging facilities, restaurants, and entertainment companies. Consumer Defensive represents companies manufacturing food, beverages, household and personal products, packaging, or tobacco. Also include companies that provide services such as education & training services. Energy represents companies that produce or refine oil and gas, oil field services and equipment companies, and pipeline operators. This sector also includes companies engaged in the mining of coal. Financial Services represents companies that provide financial services, which include banks, savings and loans, asset management companies, credit services, investment brokerage firms, and insurance companies. Health Care includes biotechnology, pharmaceuticals, research services, home healthcare, hospitals, long-term care facilities, and medical equipment and supplies. Industrials represent companies that manufacture machinery, hand-held tools, and industrial products. This sector also includes aerospace and defense companies and companies engaged in transportation and logistics services. Although the Morningstar Global Equity Classification Structure does not include a category for NGOs and similar entities, they were grouped under the category "NGO and others" due to their inability to fit into other categories for various reasons. The "NGO and others" category primarily consists of associations and unions. Technology represents companies designing, developing, and supporting computer operating systems and applications. This sector also includes companies that provide computer technology consulting services. It also includes companies manufacturing computer equipment, data storage, networking, semiconductors, and components.

The sectors with the highest percentage of corporate tweets are Energy and Industrials, with 30.6% and 21.06%, respectively. On the other hand, the Financial Services and Health Care sectors have relatively low percentages of corporate tweets and replies, with 4.66% and 2.00% of corporate tweets, respectively. Interestingly, there is a discrepancy between the number of corporate tweets and replies in some sectors. This could suggest that issues or concerns related to those industries are prompting more engagement and interaction from social media users. For instance, the Industrials sector represents a smaller proportion of corporate tweets but a more significant proportion of replies, indicating that there may be more controversial or polarizing topics related to that industry that drive discussions and debates among users. To gain more insights into the reasons for higher levels of engagement in specific sectors, further analysis of the content of the corporate tweets and replies is necessary. Such an analysis could help to shed more light on the specific issues driving the higher levels of engagement for certain sectors, as well as whether these issues are positive or negative.

The following table represents the distribution of both corporate tweets and replies across the various sectors.

Table 6 - Distribution of both corporate tweets and replies across the various sectors.

Sector	Corporate tweets	Replies
Basic Materials	2.44%	0.50%
Communication Services	1.77%	1.80%
Consumer Cyclical	19.51%	10.73%
Consumer Defensive	2.88%	1.70%
Energy	30.60%	32.59%
Financial Services	4.66%	3.81%
Health Care	2.00%	0.41%

Sector	Corporate tweets	Replies
Industrials	21.06%	43.03%
NGO and other's	10.42%	3.62%
Technology	4.66%	1.81%

The following table allows to draw conclusions regarding how a company or sector performs regarding engagement (i.e., replies) relative to their activity (i.e., corporate tweets). The maximum number of replies for a company is 1176, significantly higher than the maximum number of corporate tweets, indicating that some companies receive a lot of engagement from their audience. The standard deviation for the number of replies is much higher than for the number of corporate tweets. This indicates that there is a broader range of values for replies, which means that, once again, some companies are receiving much more engagement than others.

Table 7 - Summary of the distribution of corporate tweets and replies per company and sector.

	Corporate Tweets (Company)	Replies (Company)	Corporate Tweets (Sector)	Replies (Sector)
Min	1	1	8	29
Max	18	1176	138	3020
Mean	3.443	54.565	45.1	701.7
Std. deviation	3.259	146.523	45.945	1063.432

The following table shows companies' environmental disclosure and performance scores in different sectors on three environmental impact metrics: forest, water security, and climate change. The scores are based on a scale of 0 to 8, with eight being the best. The table also shows the standard deviation, variance, total companies, and total reported companies for each sector and metric. The table reveals that not all sectors have data for every score type, implying that some sectors may not perform well or disclose information in certain areas. Only a few sectors have data for all the metrics. These sectors are Basic Materials, Consumer Cyclical, Consumer Defensive, Energy, and Health Care. Comparing these five sectors, Consumer Defensive ranks the highest in all three metrics, while Energy ranks the lowest in all three metrics. This suggests that Consumer Defensive companies are leading in disclosing and performing well in all environmental areas. Climate change is the metric where more sectors have data. In this metric, Technology ranks the highest while Industrials ranks the lowest. Communication services, with only one company in the dataset, has no data, while the 22 NGO and other's are associations and do not report to the CDP.

Table 8 - Summary of the distribution of CDP scores across sectors.

Sector	Score type	Mean	Std. deviation	Variance	Total companies	Total reported companies
Basic Materials	Forest	1.75	3.24	10.5	5	3
	Water security	3.778	3.598	12.944	5	4
	Climate change	6.444	0.527	0.278	5	4
Communication Services	Forest	n/a	n/a	n/a	1	0
	Water security	n/a	n/a	n/a	1	0
	Climate change	n/a	n/a	n/a	1	0
Consumer Cyclical	Forest	2.958	3.123	9.755	25	18
	Water security	4.194	3.547	12.583	25	18
	Climate change	4.889	2.924	8.551	25	17
Consumer Defensive	Forest	7.077	1.32	1.744	3	3
	Water security	7.667	0.492	0.242	3	2
	Climate change	7.615	0.506	0.256	3	3
Energy	Forest	0.14	0.915	0.837	37	7
	Water security	2.212	2.888	8.339	37	18
	Climate change	4.476	3.148	9.91	37	16
Financial Services	Forest	n/a	n/a	n/a	2	0
	Water security	n/a	n/a	n/a	2	0
	Climate change	6.714	0.463	0.214	2	2
Health Care	Forest	6	0	0	2	1
	Water security	6.222	0.441	0.194	2	2
	Climate change	7.222	0.441	0.194	2	2
Industrials	Forest	n/a	n/a	n/a	30	0
	Water security	2	4	16	30	2
	Climate change	4.444	2.245	5.039	30	22
NGO and other's	Forest	n/a	n/a	n/a	22	0
	Water security	n/a	n/a	n/a	22	0
	Climate	n/a	n/a	n/a	22	0

Sector	Score type	Mean	Std. deviation	Variance	Total companies	Total reported companies
	change					
Technology	Forest	n/a	n/a	n/a	3	0
	Water security	8	0	0	3	1
	Climate change	8	0	0	3	3

Based on the following table, there appears to be an even distribution between abstract, concrete, and vague claims among the sectors analyzed, with a slight majority leaning towards vague claims. False claims were not present, apart from the Energy sector. Looking at the sectors with the most significant number of companies in the dataset (Consumer Cyclical, Energy, Industrials and NGO and other's), we can see a clear distinction between them. Consumer Cyclical and Industrials focus their communication on concrete claims, while Energy tends more towards abstract and vague, and NGO and other's mainly use vague claims. Overall, there is room for improvement in the transparency and specificity of environmental compensation claims across all sectors, with a need for more concrete and less vague claims. Fernandes et al. (2020) point to a content analysis study that revealed that 43% of ads featured vague or ambiguous claims, and 9% were classified as omissions. Despite this, a recent replication of the study found that more claims were deemed acceptable than deceptive, but misleading claims in advertising are still common. Indeed, in this dataset, abstract and vague claims account for 68.3% of all claims made. False claims were insignificant, primarily because assessing the truthfulness of a claim requires further investigation and a deep understanding of the company processes and their impact.

Table 9 -Summary of the distribution of claims.

Sector	Abstract compensation claim	Concrete compensation claim	Vague claim	False claim
Overall	33.48%	31.49%	34.81%	0.22%
Basic Materials	45.45%	36.36%	18.18%	0.00%
Communication Services	0.00%	12.50%	87.50%	0.00%
Consumer Cyclical	32.95%	42.05%	25.00%	0.00%

Sector	Abstract compensation claim	Concrete compensation claim	Vague claim	False claim
Consumer Defensive	38.46%	15.38%	38.46%	0.00%
Energy	35.51%	26.09%	37.68%	0.72%
Financial Services	42.86%	19.05%	38.10%	0.00%
Health Care	22.22%	55.56%	22.22%	0.00%
Industrials	34.74%	47.37%	17.89%	0.00%
NGO and other’s	23.40%	12.77%	63.83%	0.00%
Technology	38.10%	4.76%	57.14%	0.00%

The most interesting conclusion from the following table is that corporate praising is much more common across all sectors than consumer praising. This suggests that companies are more likely to focus on highlighting their sustainability initiatives and actions rather than promoting sustainable behavior among their consumers. Notably, the Technology sector has the highest percentage of consumer praising, indicating a potential focus on promoting sustainable behavior among customers.

Table 10 - Summary of the distribution of praising.

Sector	Corporate praising	Consumer praising
Overall	86.70%	3.55%
Basic Materials	81.82%	0.00%
Communication Services	25.00%	12.50%
Consumer Cyclical	92.05%	3.41%
Consumer Defensive	100.00%	0.00%
Energy	89.86%	1.45%
Financial Services	80.95%	4.76%
Health Care	77.78%	0.00%
Industrials	95.79%	3.16%
NGO and other’s	63.83%	6.38%
Technology	80.95%	14.29%

The following table indicates that companies' environmental actions are substantive rather than symbolic in most sectors. The overall percentage of substantive actions is 65.63%, while the percentage of symbolic actions is 24.61%. The most substantive actions were found in the Consumer

Defensive sector at 76.92%, while the lowest percentage was in the Communication Services sector at 12.5%. The highest percentage of symbolic actions was found in the Financial Services sector at 28.57%, while the lowest percentage was in the Energy sector at 21.01%. This suggests that companies generally take meaningful and concrete steps toward environmental responsibility. However, there is still room for improvement, particularly in specific sectors, such as Consumer Cyclical, Financial Services, and Industrials.

Table 11 - Summary of the distribution of actions.

Sector	Substantive action	Symbolic action
Overall	65.63%	24.61%
Basic Materials	72.73%	24.61%
Communication Services	12.50%	25.00%
Consumer Cyclical	67.05%	28.41%
Consumer Defensive	76.92%	23.08%
Energy	70.29%	21.01%
Financial Services	57.14%	28.57%
Health Care	55.56%	22.22%
Industrials	72.63%	26.32%
NGO and other’s	42.55%	27.66%
Technology	71.43%	23.81%

The Communication Services sector has the highest percentage of corporate tweets mentioning net zero or related terms (25.00%), followed by Financial Services (23.81%) and Basic Materials (18.18%). The Health Care sector has the lowest percentage of corporate tweets mentioning net zero or related terms (0.00%), followed by Consumer Cyclical (2.23%) and NGO and other’s (4.26%).

Table 12 - Summary of the distribution of net zero mentions.

Sector	Mentioned
Overall	11.31%
Basic Materials	18.18%
Communication Services	25.00%
Consumer Cyclical	2.23%
Consumer Defensive	15.38%
Energy	13.04%
Financial Services	23.81%
Health Care	0.00%

Sector	Mentioned
Industrials	17.89%
NGO and other’s	4.26%
Technology	4.76%

3.4. Modeling

In the context of natural language processing (NLP) research investigations, the modeling stage in CRISP-DM involves developing and testing NLP models to analyze text data collected in earlier stages of the research process. The aim is to extract insights and knowledge from the text data, such as sentiment analysis, topic modeling, or named entity recognition. The researcher selects the appropriate NLP models and algorithms, develops and tunes them to achieve the best possible performance, evaluates their accuracy and validity, and draws conclusions from the results. The quality of the NLP model's output depends heavily on the quality of the text data collected and the researcher's ability to select and fine-tune the appropriate models for the specific research questions.

The basic steps to perform sentiment analysis and emotion detection are as follows:

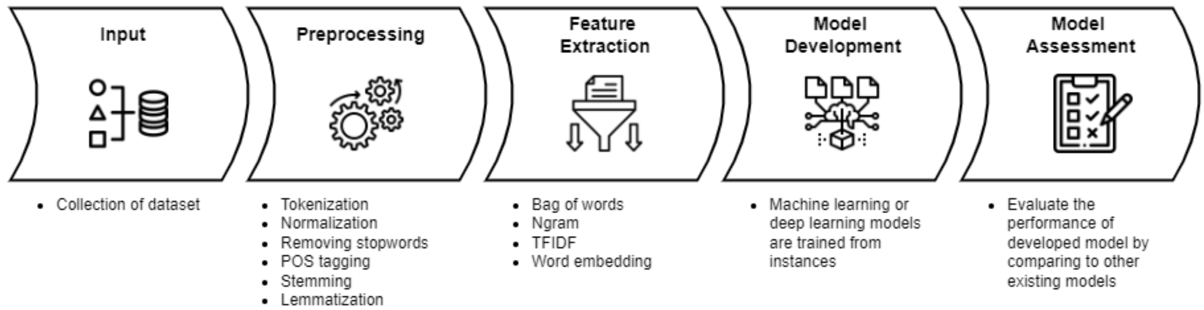

Figure 3 - Basic steps to perform sentiment analysis and emotion detection. Based on (Nandwani & Verma, 2021)

In the preprocessing stage, the general steps taken in this dissertation were: (1) normalization (normalize the text to achieve uniformity in data by converting the text into standard form, correcting the spelling of words, removing numbers and punctuation), (2) removing stop words (words that do not contribute toward sentiment analysis and emotion recognition (Tursi & Silipo, 2021)), and (3) lemmatization (involves morphological analysis to remove inflection endings from a token to turn it into the base word lemma (Tursi & Silipo, 2021). For example, “caught” is converted into “catch”). Nandwani & Verma (2021) found that removing numbers and lemmatization enhanced accuracy, whereas removing punctuation did not affect accuracy. Several techniques were implemented in the feature extraction stage to apply the techniques described in the next chapter. The techniques were (1) bag of words (a technique used in natural language processing and information retrieval to disaggregate input text into terms (Tursi & Silipo, 2021)), (2) n-gram method (term co-occurrence where the order of the words is relevant (Tursi & Silipo, 2021)) and (3) tf-idf (term frequency-inverse document frequency). The next step is model development. According to Nandwani & Verma (2021), there are five techniques for sentiment analysis and emotion detection: lexicon-based approach, machine learning-based approach, deep learning approach, hybrid approach (to overcome the drawbacks of statistical and machine learning approaches), and transfer learning approach (the use of pre-trained models in other similar domains). The approach adopted in this dissertation is the transfer learning approach, which can save time and produce more efficient results than the rest (Nandwani & Verma, 2021).

Two models were used, one for sentiment analysis and one for emotion detection, as they each proved to be the most accurate in their respective tasks. The models used in this dissertation were BERT (Vaswani et al., 2017) and BERTweet (Quoc Nguyen et al., 2020). BERT, an acronym for Bidirectional Encoder Representations with Transformers, represents a Deep Neural Network (DNN) architecture incorporating the latest advancements in Deep Learning, specifically in Natural Language Processing (NLP). Developed and introduced by Google in 2018, BERT has exhibited State-of-the-Art (SOTA) performance across various benchmarks in Natural Language Understanding (NLU). At its core, BERT operates on the foundational concept of acquiring a language model called Encoder Representation, which enables the prediction of tokens or sequences within a given contextual framework. In practical terms, when discussing BERT, we typically refer to the BERT language model itself. The intrinsic power of pre-trained Deep Learning Language Models lies in their ability to represent language dynamics, thereby facilitating downstream tasks comprehensively. For example, by appending a final layer to the pre-trained language model and fine-tuning its last layer(s) for a specific task, such as classification, superior results can be achieved with reduced reliance on annotated data. The BERT architecture revolves around the application of Transformers, a fundamental component that harnesses the capability to predict a token by intelligently considering each token present in the sequence. This transformative concept addresses the critical issue of long-distance dependencies within language, where interpreting a single word may necessitate consideration of preceding or subsequent terms in the sequence. Consequently, BERT's bidirectional approach efficiently tackles these dependencies, amplifying its effectiveness in language understanding tasks.

3.5. Evaluation

In NLP research investigations following the CRISP-DM methodology, the evaluation stage is a crucial step that involves measuring the performance of the NLP models on specific tasks using metrics such as precision, recall, F1 score, and accuracy. The goal of the evaluation stage is to assess the effectiveness of the models and determine if they meet the project's objectives. The datasets used for sentiment analysis and emotion detection are presented in Table 13. For sentiment analysis, the models presented in Table 14 were tested. After an extensive trial and error period, all datasets regarding sentiment analysis were used. It was found that using all datasets resulted in greater accuracy than just using one or a combination of datasets.

Table 13 - Datasets for sentiment analysis and emotion detection.

Dataset	Size	Sentiment/emotion analysis	Sentiment/emotions	Domain
Archeage	1,718	Sentiment analysis	Positive and negative	Games
OMD	1,906	Sentiment analysis	Positive and negative	Presidential Debate
HCR	1,908	Sentiment analysis	Positive and negative	Health Care Reform
movie	561	Sentiment analysis	Positive and negative	Movies
Narr	1,227	Sentiment analysis	Positive and negative	Generic
aisopos	278	Sentiment analysis	Positive and negative	Generic
sanders	1,224	Sentiment analysis	Positive and negative	Business
SemEval16	12,216	Sentiment analysis	Positive and negative	Generic
SemEval18	1,859	Sentiment analysis	Positive and negative	Equity Evaluation Corpus
sentiment140	359	Sentiment analysis	Positive and negative	Generic
SentiStrength	2,289	Sentiment analysis	Positive and negative	Generic
STS-gold	2,034	Sentiment analysis	Positive and negative	Generic
Vader	4,196	Sentiment analysis	Positive and negative	Generic
SemEval-2018-Ec-En	7,724	Emotion detection	Anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust	Generic
SemEval-2018-4 labels	5,051	Emotion detection	Anger, joy, optimism, sadness	Generic
SemEval-2018-Irony detection	4,601	Emotion detection	Irony and non-irony	Generic
SemEval-2017-Sentiment analysis	59,899	Sentiment analysis	Positive, negative, and neutral	Generic
SemEval-2018-El-reg-En[5]	4,300	Emotion detection	Anger, fear, joy, and sadness	Generic
SemEval-2018-	3,268	Emotion detection	Anger, fear, joy, and	Generic

[5] This dataset is generally used for the detection of emotion intensity in regression tasks (ranging between 0 being the lowest amount of emotion and 1 being the highest amount of emotion). Only values above 0.50 were included.

Dataset	Size	Sentiment/emotion analysis	Sentiment/emotions	Domain
El-oc-En[6]			sadness	
GoEmotions[7]	57,730	Emotion detection	Admiration, amusement, approval, caring, desire, excitement, gratitude, joy, love, optimism, pride, relief, anger, annoyance, disappointment, disgust, embarrassment, fear, grief, nervousness, remorse, sadness, confusion, curiosity, realization, surprise	Generic

Table 14 - Sentiment Analysis Models

Model	Recall	Precision	Specificity	F-measure	Accuracy
bert_en_cased_L-24_H-1024_A-16 2 epochs	0.7777	0.667	0.8892	0.7669	0.7766
bert_en_cased_L-24_H-1024_A-16 3 epochs	0.7651	0.7731	0.885	0.7681	0.7785
bertweet_base 2 epochs	0.7817	0.7974	0.8936	0.7856	0.7904
bertweet_base 3 epochs	0.7953	0.7972	0.8996	0.7945	0.7889

Overall, the models achieved relatively high accuracy scores, ranging from 0.7766 to 0.7904. The bertweet_base (2 epochs) model achieved the highest accuracy score and was the model chosen for sentiment analysis.

Recall is a measure of completeness, the number of correctly classified instances concerning the total number of objects belonging to that class (Berthold et al., 2020). Precision can be seen as a measure of exactness. The precision for a class is the number of instances correctly labeled as belonging to that class for the total number of elements labeled as belonging to that class (Berthold et al., 2020). Specificity measures the model's capability of recognizing what does not belong to the

[6] This dataset is generally used for the detection of emotion intensity in ordinal classification (0: no emotion can be inferred to 4: high amount of emotion can be inferred). Only moderate and high intensity were selected.
[7] https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html

positive class. If the model recognizes what does not belong to that class, the result is 0 “false positives,” which means no data objects are misclassified into the positive class. Specificity is then 1 (Berthold et al., 2020). F-measure can be interpreted as a weighted average of precision and recall, where it reaches its best value at one and its worst score at zero. Lastly, accuracy is a measure of the average correctness of the classifier and is calculated across all classes. An accuracy of 1 means that the classified values are the same as the original class values (Berthold et al., 2020). The procedure for evaluating a classification model involves dividing the dataset into training and test sets, with a proportion of 70% / 30%, respectively (Quinn, 2020).

For emotion detection, several combinations of models and datasets were assessed. As Table 13 shows, different datasets contain different emotions, some containing multiple emotions for the same sentence, turning this into a multi-label classification problem. Ultimately, the choice was between model accuracy and emotion diversity. For example, using the BERTweet model with the SemEval-2018-El-reg-En and SemEval-2018-El-oc-En proved to be the most accurate with a recall of 0.8138, precision of 0.8352 and accuracy of 0.8426. The problem with this implementation was that these two datasets only covered anger, fear, joy, and sadness. The dataset used was Go Emotion but mapped[8] to the six basic emotions proposed by Ekman: anger, disgust, joy, fear, sadness, and surprise. Table 15 presents performance metrics for several models that have been evaluated. The models are listed in the first column, and the remaining columns show various metrics used to assess their performance. The metrics used are recall, precision, f_0.5, f_1, and f_2, which measure how well the model can classify examples correctly. f_0.5, f_1, and f_2 are all variations of the F-score metric (harmonic mean of precision and recall). Their values represent the weight given to either precision or recall. The f_0.5 score gives more weight to precision than recall, meaning that the model is penalized more for false positives than false negatives. The f_2 score gives more weight to recall, meaning that the model is penalized more for false negatives than false positives. The table also includes percentages for an exact match, partial match, miss, empty, and non-empty. Exact match measures how often the model's prediction is the same as the ground truth, while partial match measures how often the model's prediction overlaps with the ground truth. The miss percentage shows how often the model fails to make a prediction, while the empty percentage shows how often the model predicts a "no answer" label when there is an answer. Finally, the non-empty percentage shows how often the model correctly predicts a non-empty answer. The model chosen was bert_en_wwm_cased_L-24_H-1024_A-16 (3 epochs). The parameters are the following: L - number of transformer blocks; H - hidden layers size; A - number of attention heads.

[8] https://github.com/google-research/google-research/blob/master/goemotions/data/ekman_mapping.json

Table 15 - Emotion Detection Models

Model	Recall	Precision	f_0.5	f_1	f_2	Exact match (%)	Partial match (%)	Miss (%)	Empty (%)	Non-empty (%)
bert_en_cased_L-24_H-1024_A-16 2 epochs	0.85	0.67	0.7	0.75	0.81	55.8	33.5	10.7	10.7	89.3
bert_en_cased_L-24_H-1024_A-16 3 epochs	0.87	0.7	0.73	0.78	0.83	58.3	32.5	9.2	9.2	90.8
bert_en_wwm_cased_L-24_H-1024_A-16 2 epochs	0.87	0.7	0.73	0.78	0.83	58.7	32.1	9.2	9.2	90.8
bert_en_wwm_cased_L-24_H-1024_A-16 3 epochs	0.88	0.71	0.74	0.78	0.84	59.3	31.9	8.7	8.7	91.3
bertweet_base 2 epochs	0.87	0.69	0.72	0.77	0.83	57.84	32.85	9.31	8.7	91.3
bertweet_base 3 epochs	0.7	0.87	0.73	0.78	0.83	57.9	32.9	9.2	9.2	90.8

3.6. Deployment

In the CRISP-DM methodology, the deployment stage is the final step in the data mining process. In the present case, the implementation of the NLP models culminated in the writing and presentation of the research results, i.e., the preparation of the dissertation document and its presentation. This is a crucial step in communicating the findings of the research investigation to the scientific community and contributing to the field of NLP. The content of the deployment stage is present in the following two chapters of the dissertation, including the conclusion, business and scientific contributions, limitations, recommendations, and next steps.

CHAPTER 4

Results and Discussion

This chapter presents the results and discusses the findings, addressing the three key research questions. Each chapter segment corresponds to one specific research question, enabling a comprehensive exploration of the topic at hand.

4.1. Consumer response analysis

The first research question focuses on determining the extent to which consumers' responses to corporate tweets are related to greenwashing, climate issues, or something else.

The first step was to conduct topic modeling LDA, a generative unsupervised probabilistic algorithm that finds the top K topics in a dataset described by the most relevant N keywords. In other words, the documents in the dataset are represented as random mixtures of latent topics, where a Dirichlet distribution over a fixed vocabulary characterizes each topic. “Latent” means we must infer the topics rather than directly observe them. To determine the optimal number of topics, the elbow method was used, which involved running the k-means algorithm on the input data for a range of values of k, calculating the within-cluster sum of squared errors (SSE) for each k value, and plotting the SSE values in a scatter chart to identify the elbow point (Tursi & Silipo, 2021).

Three topics were identified. The first topic (orange) seems to be focused on environmental issues related to fuel, climate, and carbon emissions. The second topic (blue) appears to be about supporting specific causes or initiatives, such as stopping certain activities or promoting alternative energy sources. The third topic (green) seems to be about customer service issues related to flight and airline experiences, such as requesting refunds or addressing problems with time and money.

Figure 4 - Topic modeling LDA Replies

The following analysis was document frequency, based on the principle that not all words carry the same information. More frequent words, for example, clearly indicate the text topic. The method used to represent words in the tag cloud is called TF-IDF or term frequency-inverse document frequency. It measures the importance of a word in a document or a corpus of documents. The method is based on two measures: term frequency (TF) and inverse document frequency (IDF). Term frequency measures how often a word appears in a document. The more often a word appears in a document, the higher its term frequency score. Inverse document frequency measures how rare a word is in a corpus of documents. The rarer a word is in a corpus, the higher its IDF score. IDF is calculated by taking the logarithm of the ratio of the total number of documents in the corpus to the number of documents containing the word. The resulting IDF score is then multiplied by the absolute frequency score to obtain the TF-IDF score for each word. The TF-IDF score is higher for words that occur frequently in a document (reply) but rarely in the corpus (all replies) and lower for words that occur frequently both in the document and in the corpus. The lowest TF-IDF score corresponds to words in all the corpus documents. The TF-IDF method is commonly used in information retrieval and text categorization models to identify the most relevant words in a document or a corpus. The resulting TF-IDF scores are used to generate the tag cloud, where the size of each word in the cloud corresponds to its TF-IDF score (Tursi & Silipo, 2021).

Figure 5 - TF-IDF Replies

Next was the N-Grams analysis. N-grams are sequences of N neighboring items, such as phonemes, syllables, letters, words, or DNA bases. Unlike word pairs, the order of words in N-grams is important, so T1 followed by T2 is a different bi-gram from T2 followed by T1 (Tursi & Silipo, 2021).

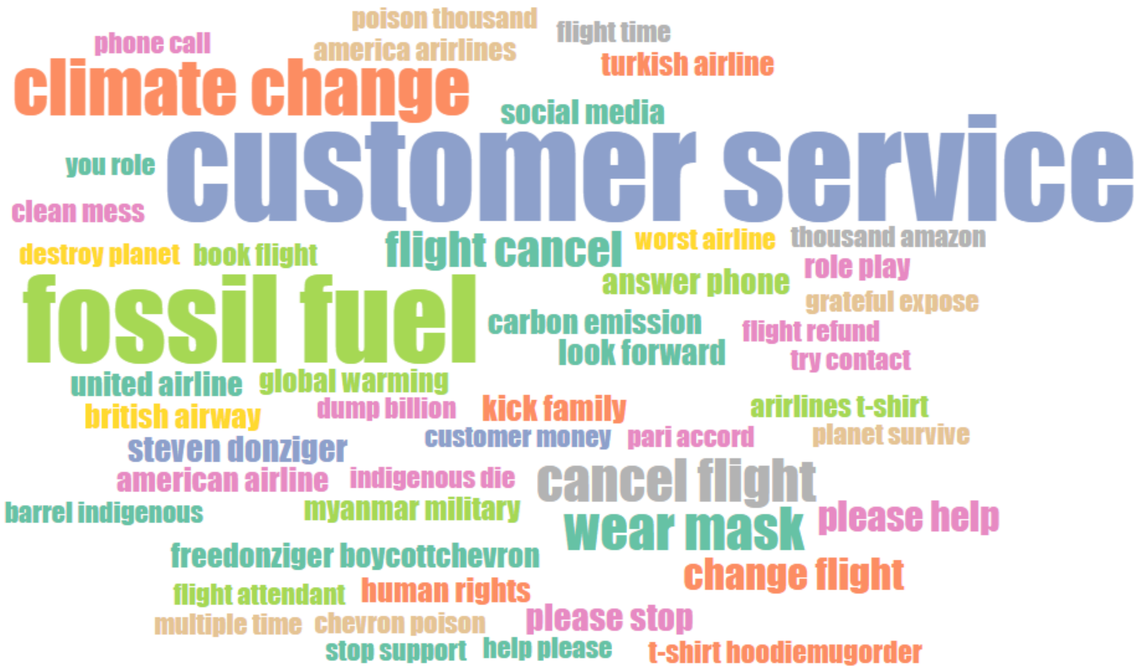

Figure 6 - Bi-grams Replies

Figure 7 - Trigrams Replies

Lastly, a dictionary was created using standard climate change and greenwashing terms to determine how consumers' responses to corporate tweets are related to greenwashing, climate issues, or something else. Of the 7017 replies, 13.9% mention climate change or greenwashing terms, meaning climate might not be a primary consumer concern. As the literature reveals, consumers prioritize price over the company's respect for the environment and climate change when making a purchase (Beriain Bañares et al., 2021), suggesting that although consumers have evolved positively in terms of pro-environmental knowledge, awareness, attitudes, and behaviors, not every consumer is willing to pay extra for environmentally friendly products or services (ShabbirHusain & Varshney, 2019). As the tag clouds demonstrate, although some climate-related terms are found, mainly the replies refer to customer service, money, and service complaints.

The following graphs present notable insights into the climate response for their respective variables.

The sector is one variable where there is a significant difference between members. Energy and Financial Services elicit a much stronger climate response than the remaining sectors. Energy had 2287 replies in the dataset, 931 mention climate-related terms. Financial Services had 267 replies in the dataset, 171 mentioned climate-related terms.

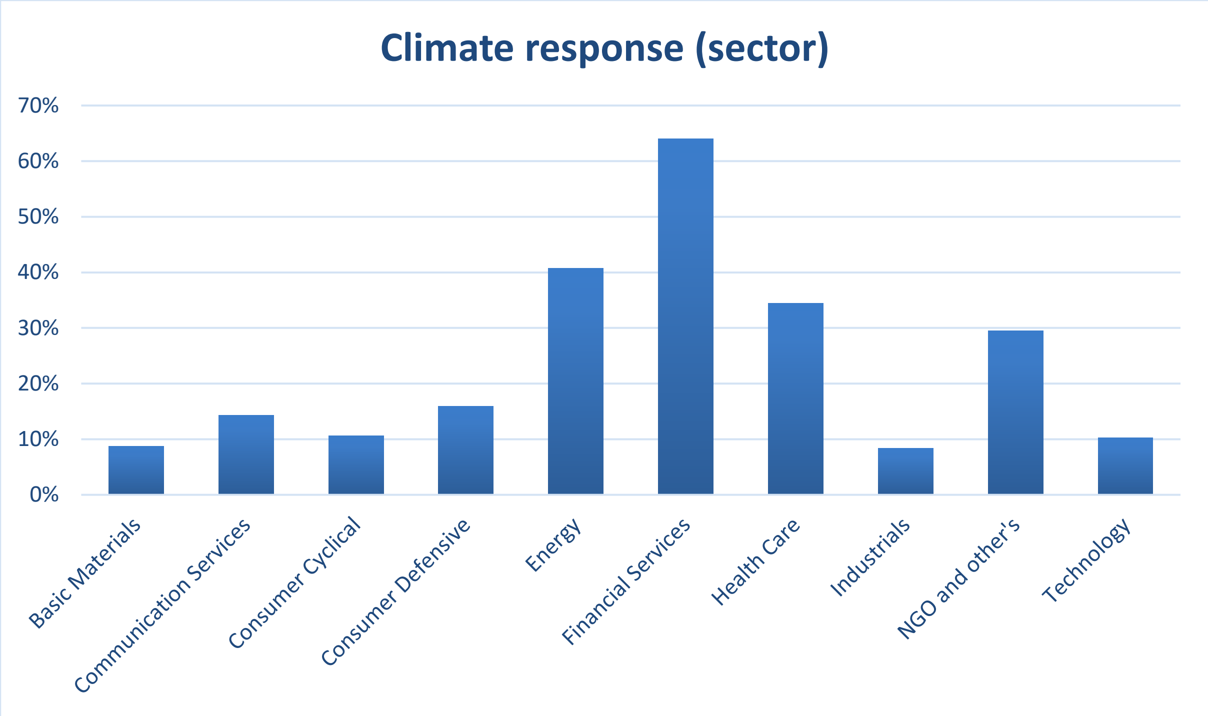

Figure 8 - Sector climate response

As we can see, having a CDP climate change score of zero also elicits a much stronger climate response. The dataset has 14 companies with a score of zero, 64 corporate tweets, and 1396 replies, 493 of which are related to climate issues.

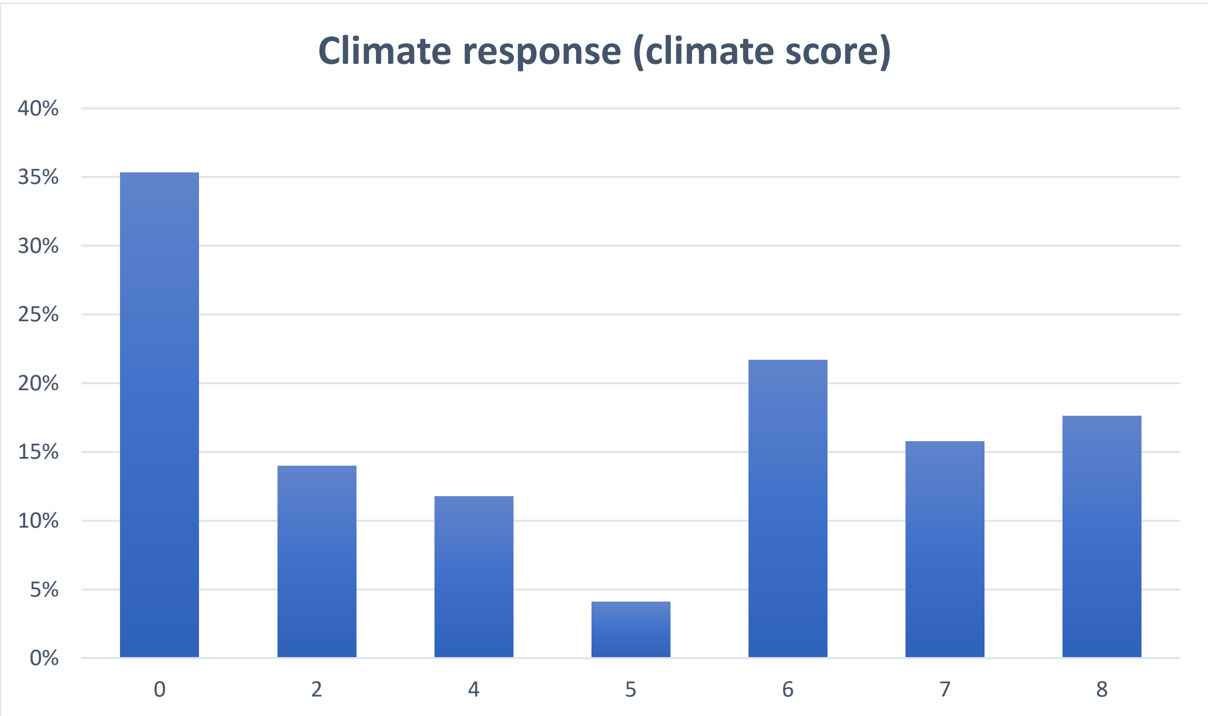

Figure 9 - Climate score response

4.2. Consumer sentiment analysis

This research question examines the feelings consumers demonstrate towards campings and determines if feelings depend on the type of claims, actions, praising, sector, and climate leadership score.

Sentiment analysis and emotion detection were conducted. The following charts present the distribution of sentiments and emotions for all replies. The overall sentiment distribution is heavily skewed towards negative sentiments, with 68.41% of replies being negative, 26.24% being positive, and only 0.04% being neutral. This suggests that most of the replies express dissatisfaction or criticism. Regarding emotions, the data suggests that anger, joy, and surprise are the most prevalent emotions, being present in 24% to 26% of replies. The emotion literature has been somewhat unclear about the affective valence of surprise. Surprise has been depicted as a pre-affective state or as an emotion that can be both positive and negative, depending on the goal relevance of the surprising event. Noordewier & Breugelmans (2013) concluded that surprise has a negative valence but is not as strong as sadness or fear. It is now possible to conclude that negative emotions also dominate the emotional state of replies.

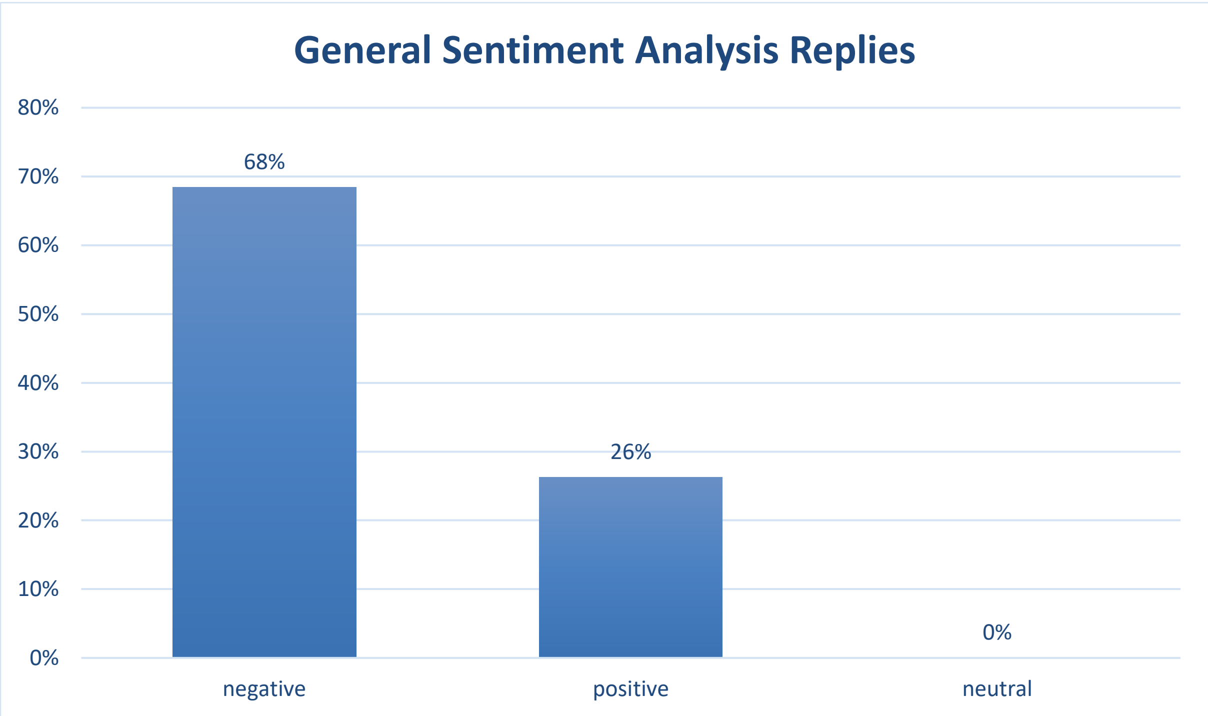

Figure 10 - Sentiment distribution

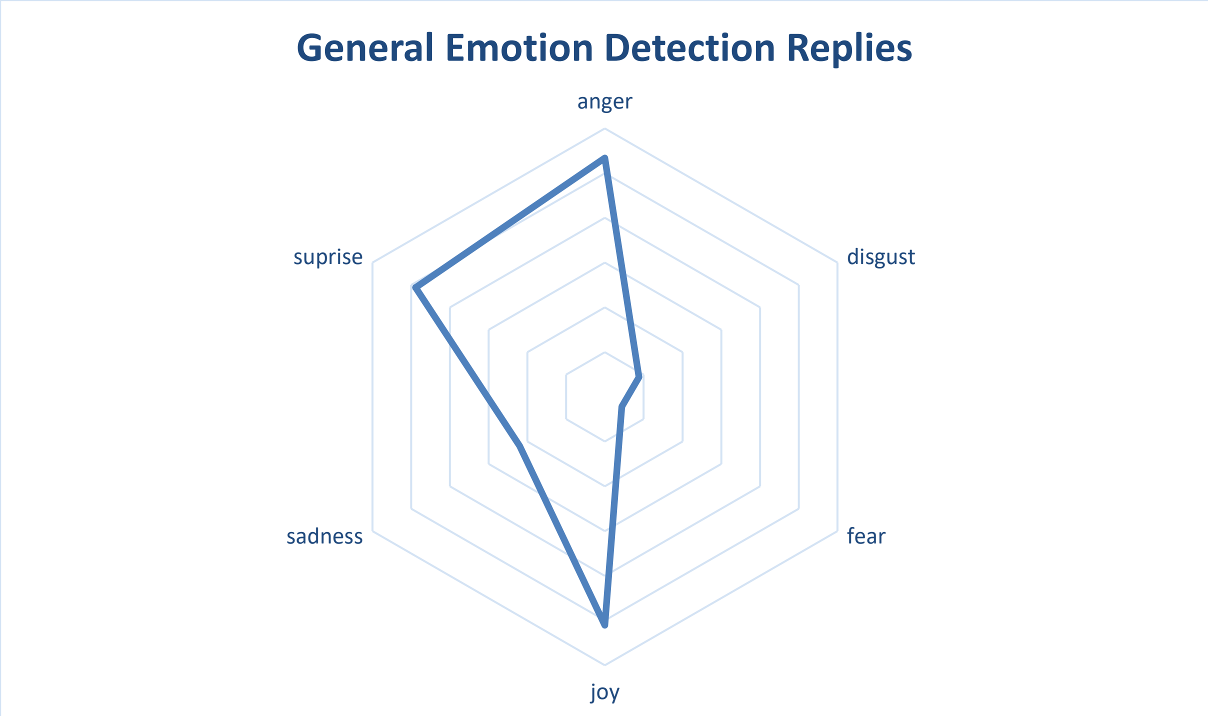

Figure 11 - Emotion distribution

In terms of claims (abstract, concrete, and vague claims), we can see that there is no significant difference between them. Concrete compensation claims generate the most negative emotions (74.56%), while abstract compensation claims generate the least (70.27%). It is possible to see from the radar charts that the type of claim has some effect on the emotions expressed by consumers, especially in terms of anger. Abstract and vague claims generate more angry emotions than concrete claims. It is important to note that different authors diverge in their findings regarding the impact of claim types. Some assert that abstract, vague, and false claims contribute to higher levels of perceived greenwashing (Neureiter & Matthes, 2022), while others argue that vague claims in green advertising are unrelated to perceived greenwashing and do not enhance consumers skepticism compared to non-deceptive claims (Schmuck et al., 2018).

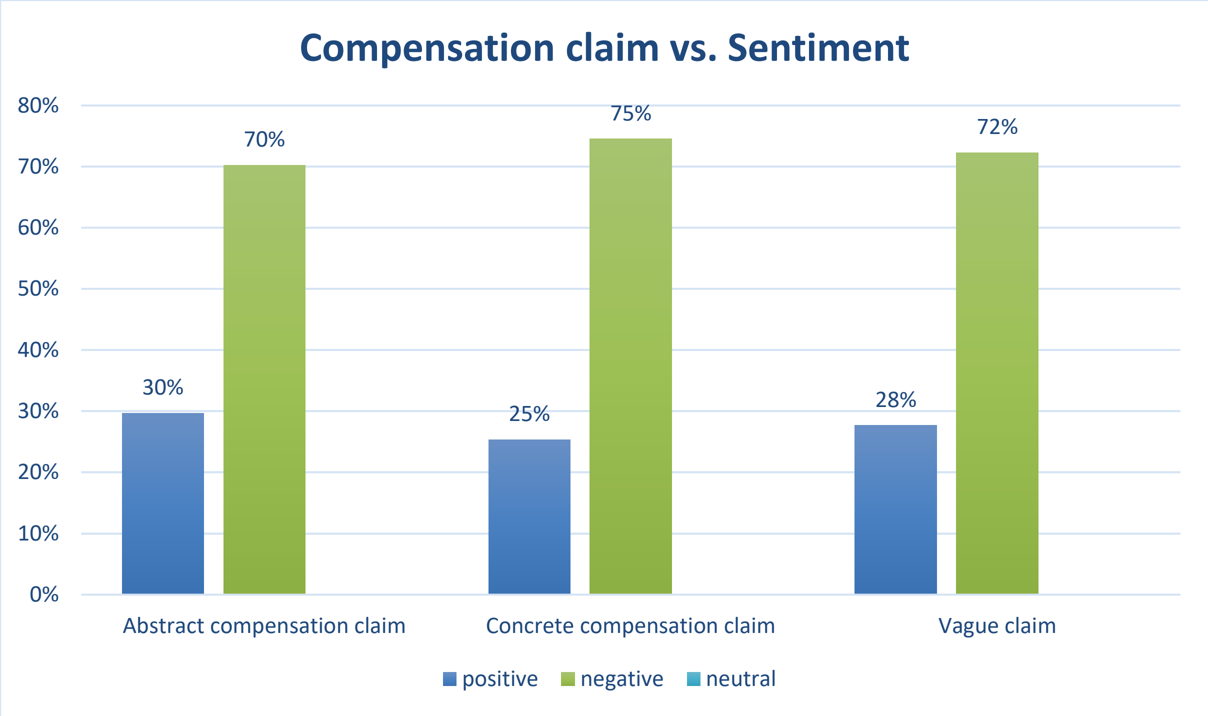

Figure 12 - Claim vs. Sentiment

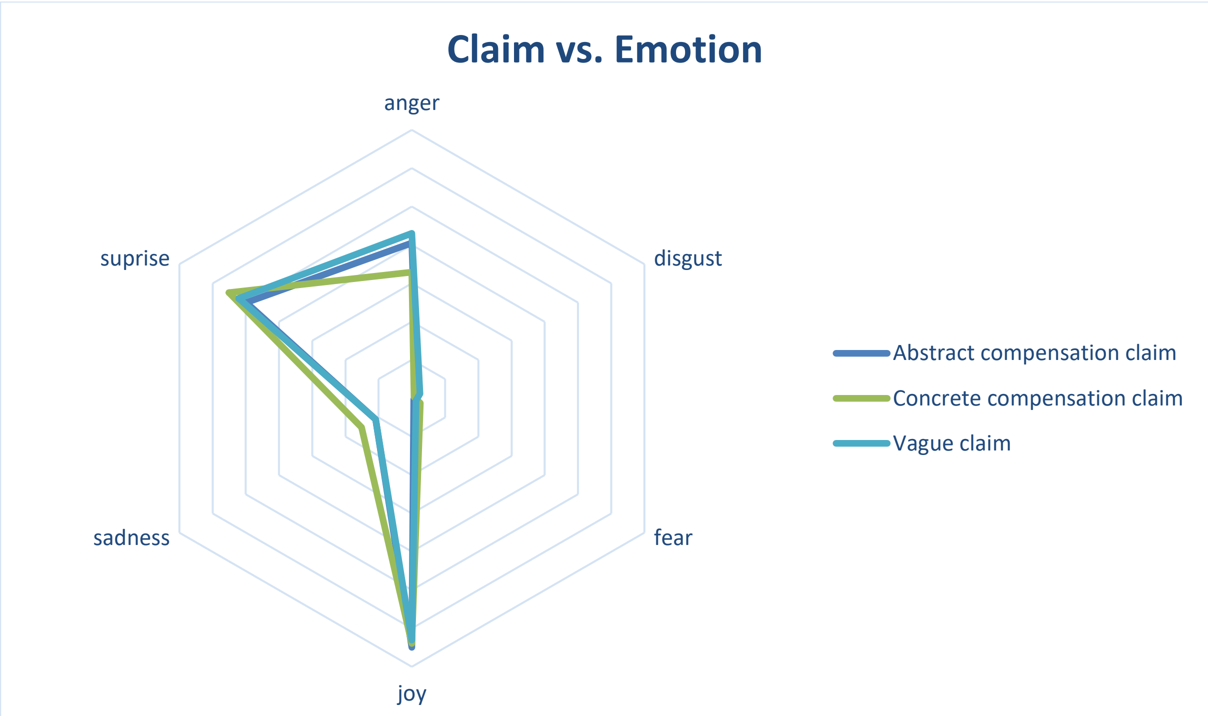

Figure 13 - Claim vs. Emotion

Actions (substantive and symbolic actions) also have very similar sentiment and emotion profiles. Symbolic action generates slightly more negative comments, and substantive actions elicit slightly more surprise. Zhang and Sun (2021) note that consumers generally respond more positively to companies' substantive environmental actions than symbolic ones.

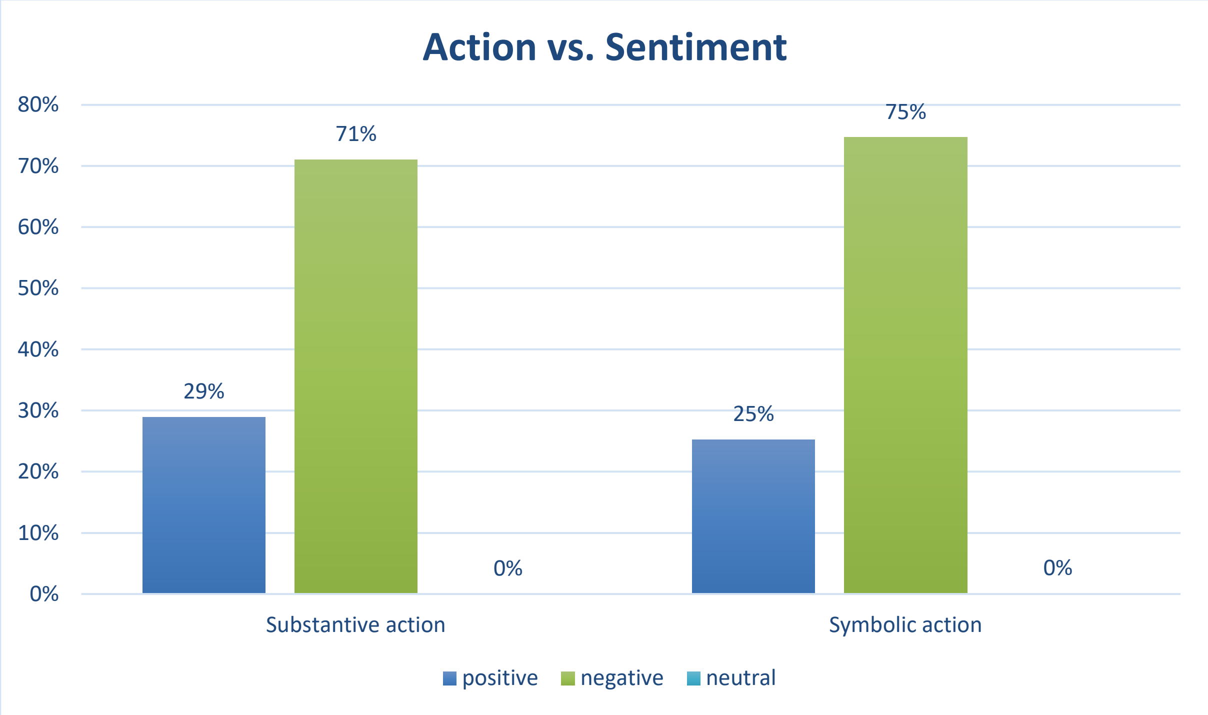

Figure 14 - Action vs. Sentiment

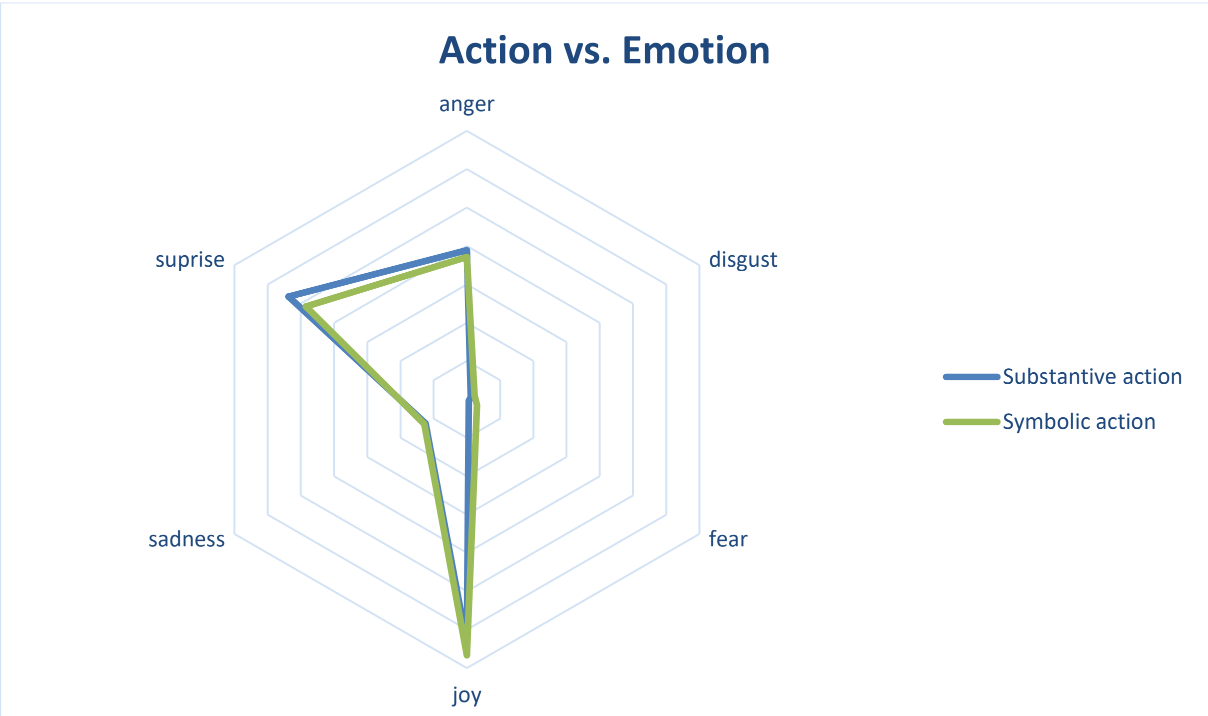

Figure 15 - Action vs. Emotion

Sector presents the most significant differences both in sentiment and emotions. Regarding sentiment, Communication Services, Energy, Financial Services, Industrials and NGO and other's have significantly more negative sentiment than the other five sectors. Basic Materials and Technology have significant positive sentiment, Technology being the only sector where positive sentiment is superior to negative. The emotion profile for the sectors also presents some interesting findings. Communication Services, Consumer Defensive, Energy and NGO and other's show significantly more anger than the remaining sectors. Health Care has a very significant value for joy. In their research, Neureiter & Matthes (2022) point out that consumers perceive green airline ads as incongruent with their existing mental representation of flying being environmentally harmful, leading to increased cognitive attention and perceptions of greenwashing. The same could be said for the Energy and Industrial sectors.

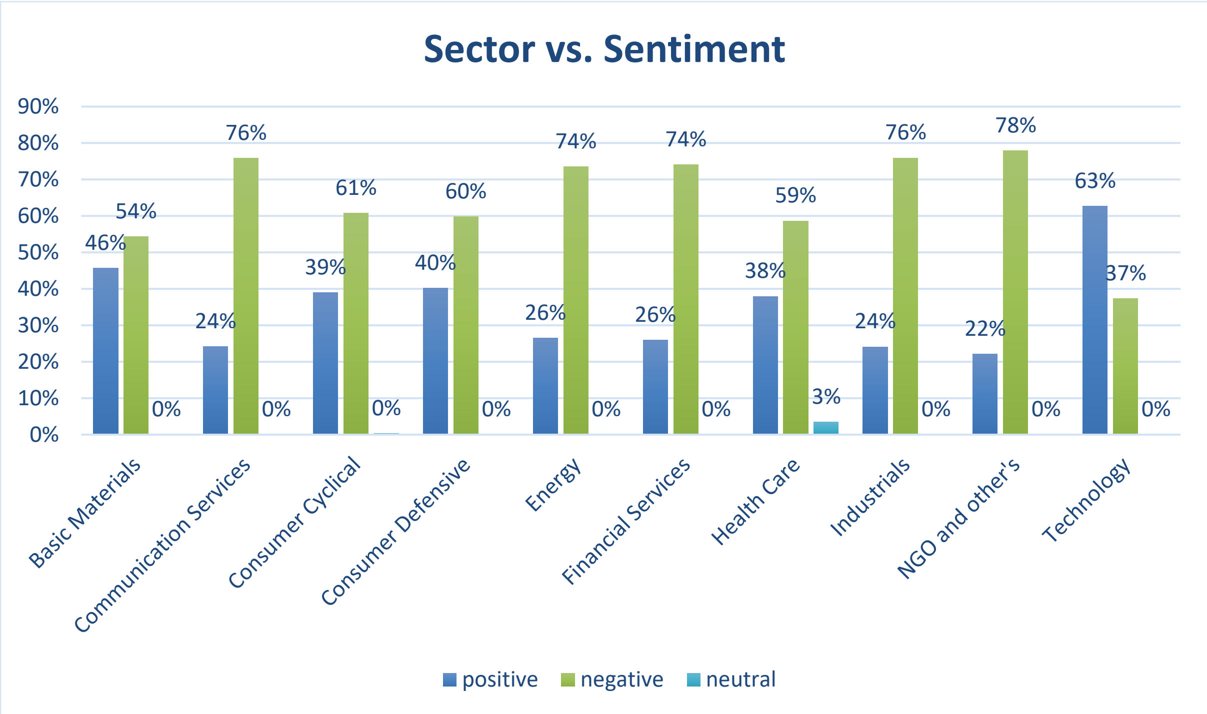

Figure 16 - Sector vs. Sentiment

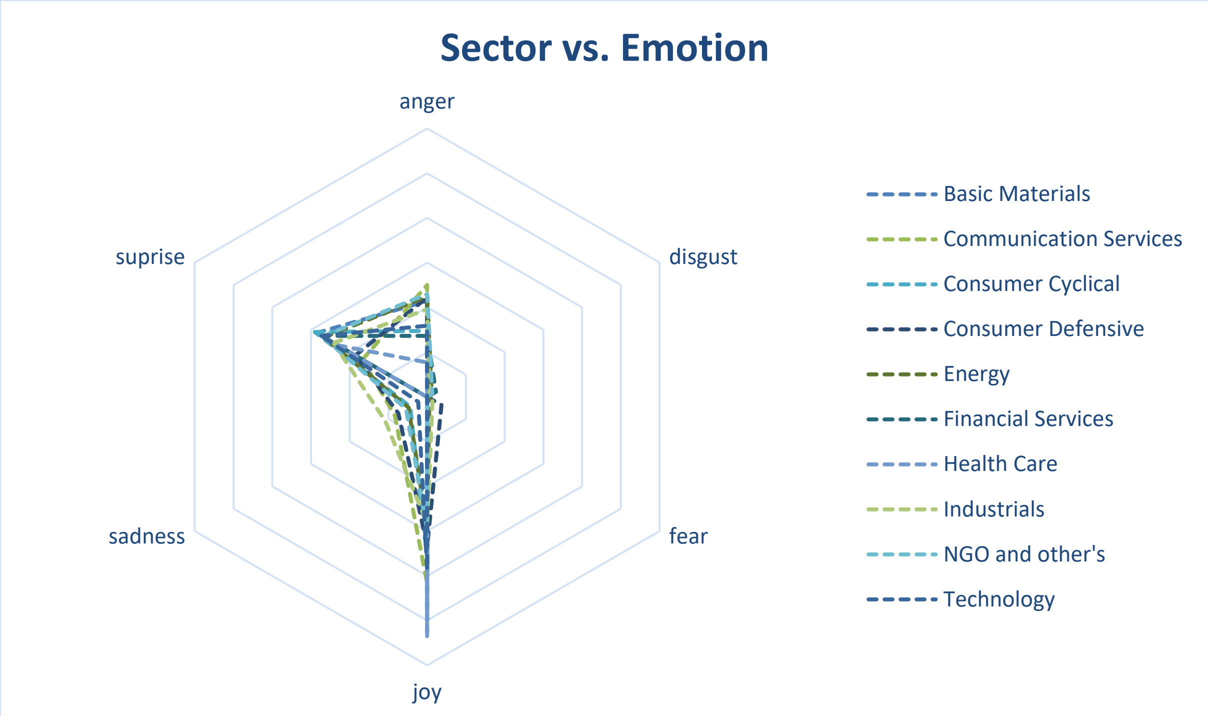

Figure 17 - Sector vs. Emotion

Praising also shows some differences. While consumer praising generates 39.3% of positive comments, corporate praising only generates 27.7%. This may suggest that praising consumers for their sustainable behavior or purchases might be a more exciting course of action than praising the company for its social actions, such as its sustainability initiatives, to avoid backlash. Regarding emotions, it is worth mentioning that corporate praising generates slightly more anger, and consumer praising generates significantly more joy. Researchers Christis and Wang (2021) highlight that participants exposed to consumer praise tended to trust the brand more.

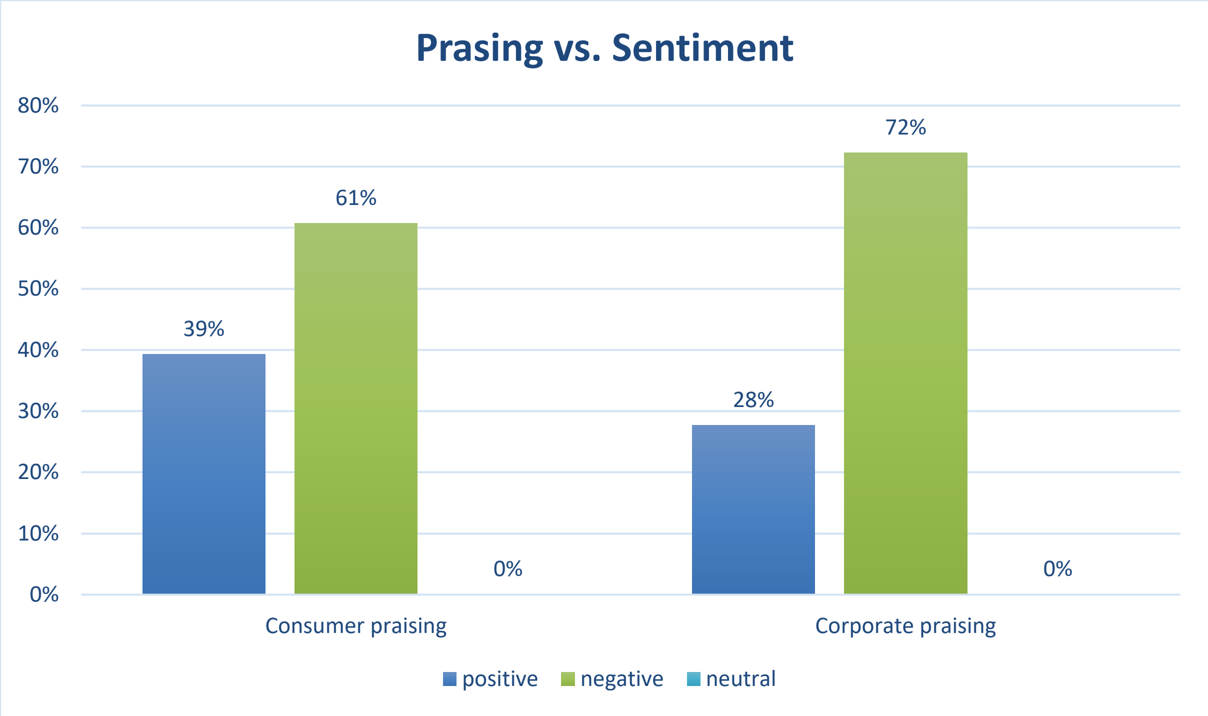

Figure 18 - Praising vs. Sentiment

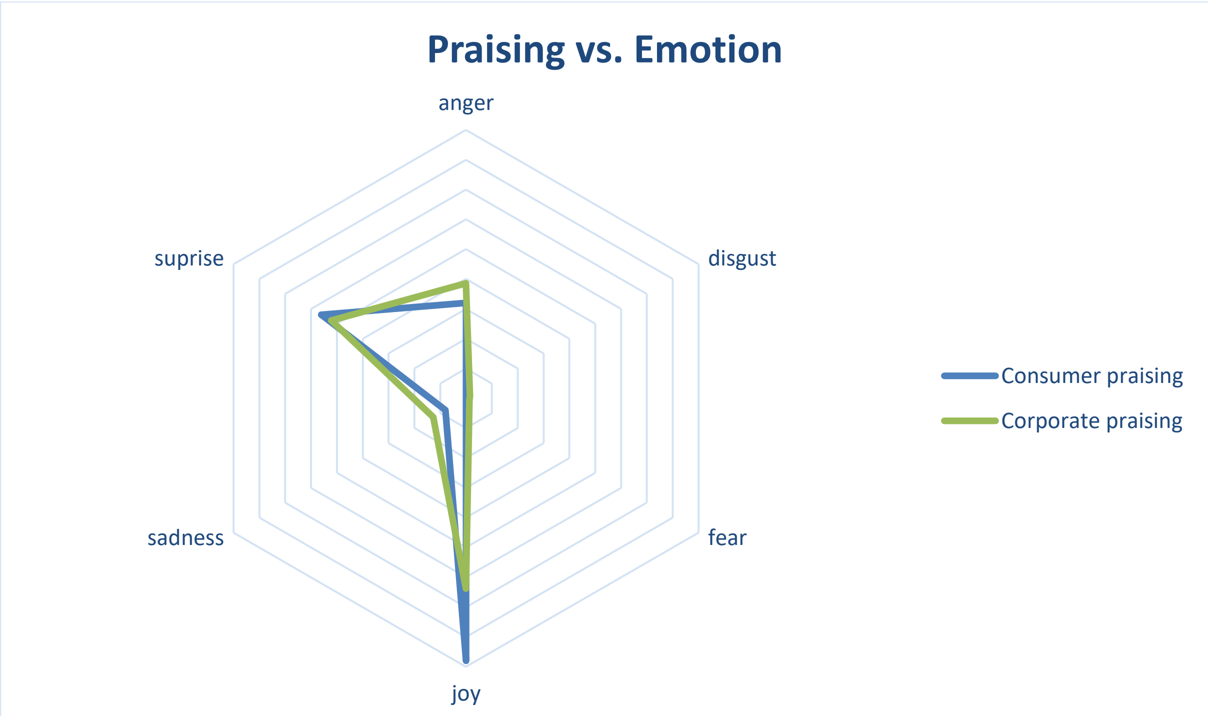

Figure 19 - Praising vs. Emotion

The presence of "net zero" or related terms did not statistically affect consumer reaction to corporate tweets. However, a slight difference in sentiment was observed, with the absence of these terms generating slightly more anger and the presence of the terms generating slightly more sadness, meaning that consumers were slightly more likely to feel angry when the terms were absent and slightly more likely to feel sad when the terms were present.

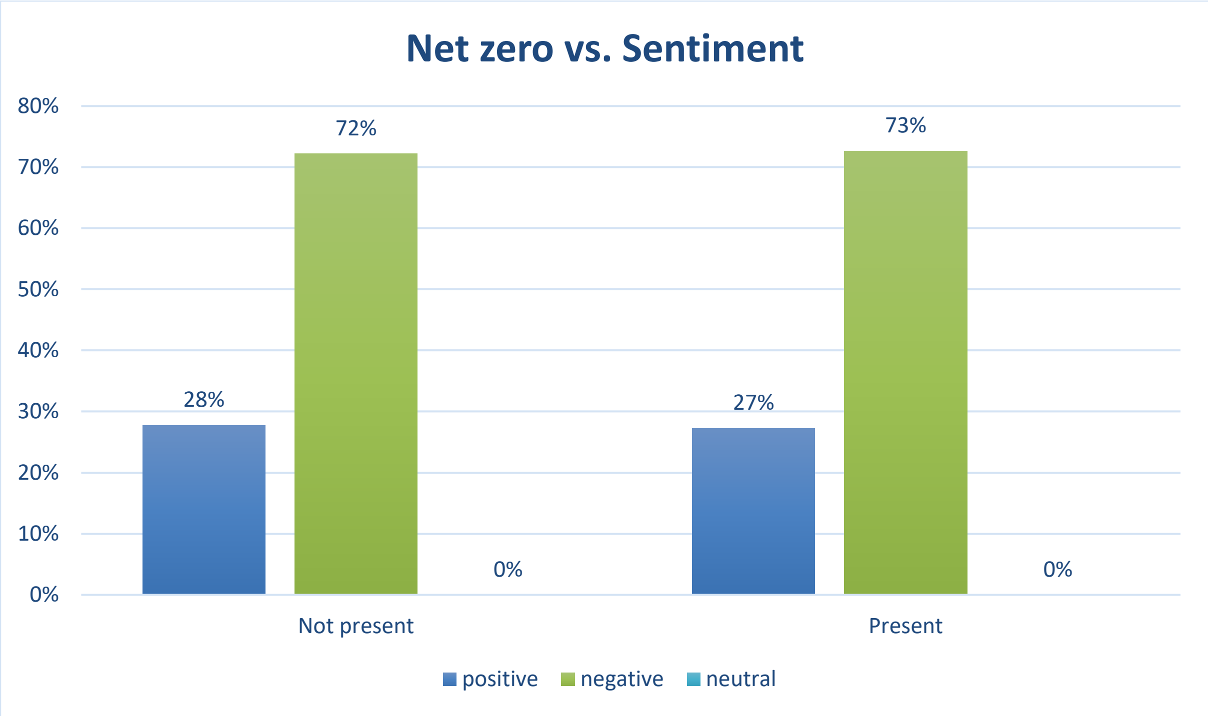

Figure 20 - Net-zero vs. Sentiment

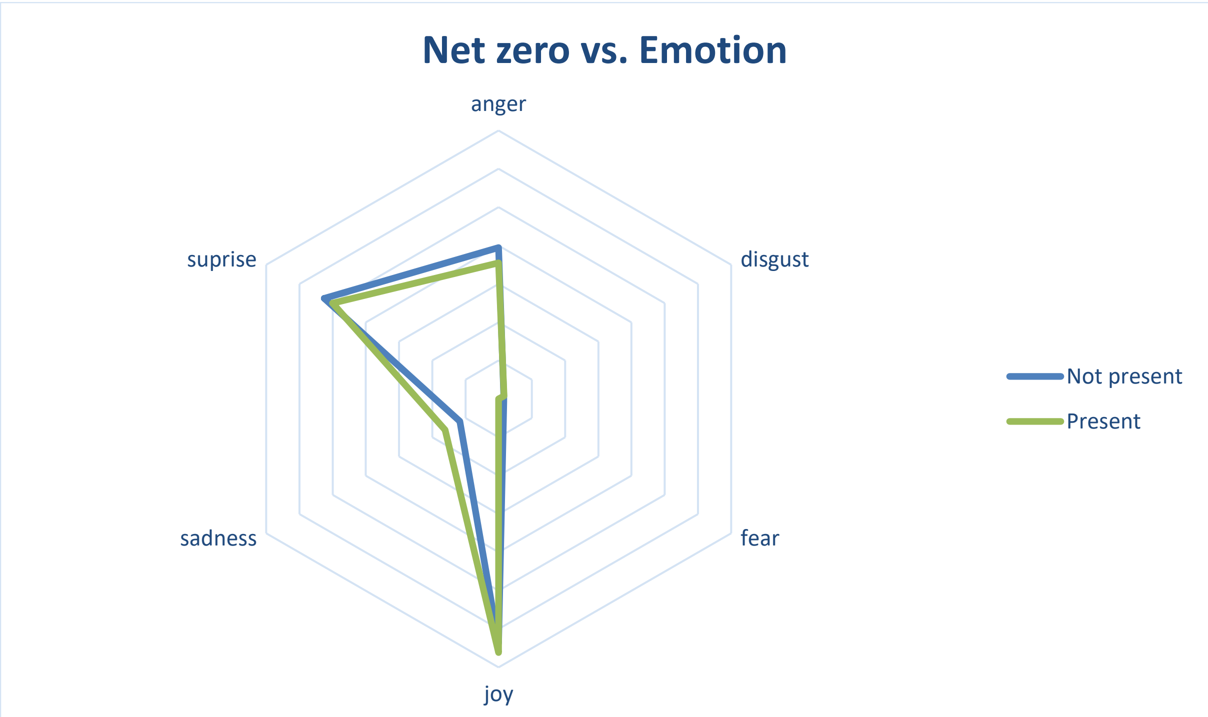

Figure 21 - Net-zero vs. Emotion

Climate score has a significant effect on consumer response. A weak positive correlation was found between climate score and negative sentiment, indicating that an increase in climate score is associated with a slight increase in negative sentiment. The linear trendline supports this unexpected finding. The emotion detection analysis found the most significant differences in consumer sentiment. Anger was found to decrease significantly as climate score increased, while sadness and joy were found to increase significantly.

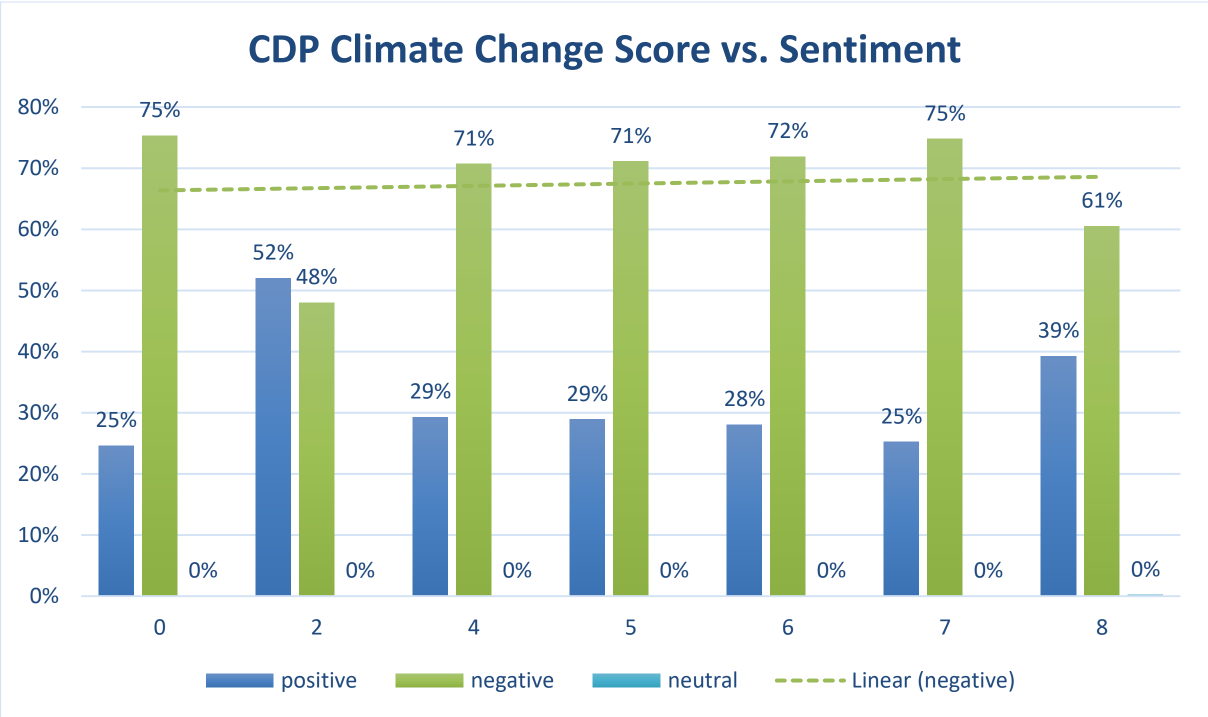

Figure 22 - CDP Climate Change Score vs. Sentiment

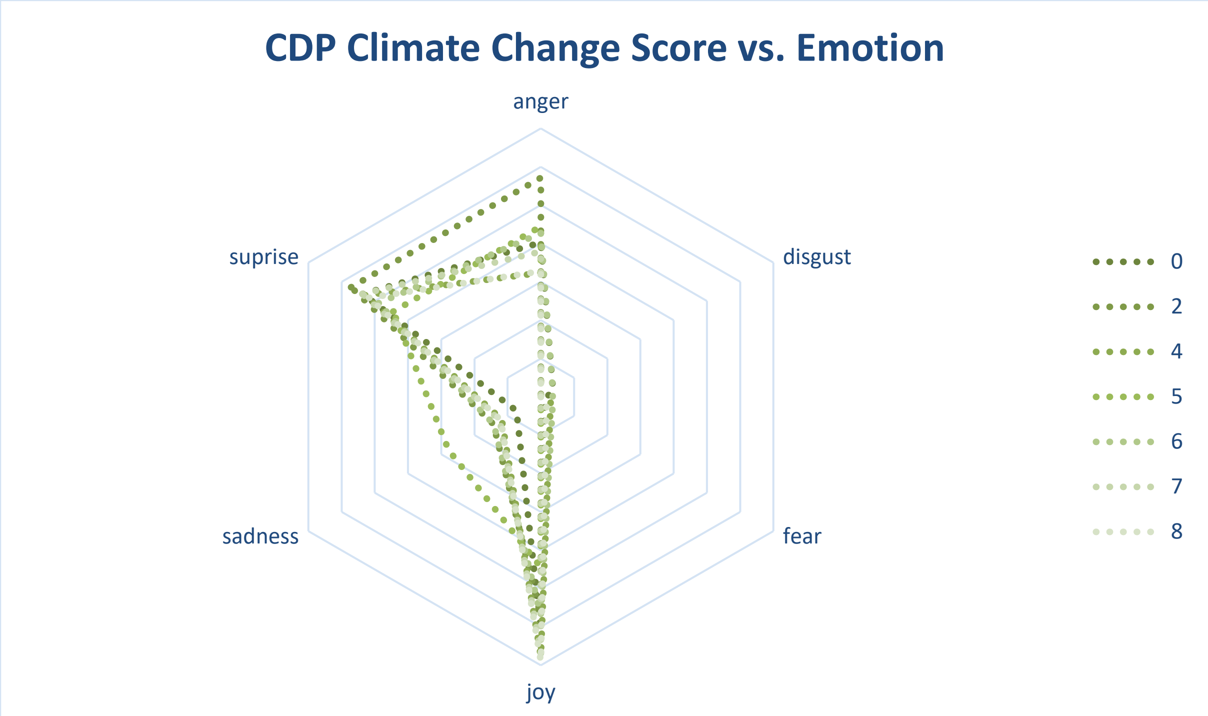

Figure 23 - CDP Climate Change Score vs. Emotion

The following analysis consists of implementing a logistic regression. Logistic regression predicts the probability that an observation falls into two categories of dichotomous dependent variables based on one or more independent variables that can be continuous or categorical. There are multiple assumptions to perform binary logistic regression: (1) the dependent variables should be measured on dichotomous scales; (2) have one or more independent variables, which can be either continuous or categorical; (3) have independence of observations and the dependent variable should have mutually exclusive and exhaustive categories, and (4) there needs to be a linear relationship between any continuous independent variables and the logit transformation of the dependent variables.

A logistic regression was performed to ascertain the effect of climate score, presence of net zero terms, company sector, type of claim, action, and praising on the likelihood that the corporate tweet would generate a negative response. The logistic regression model was statistically significant, $\chi2(12) = 194.203$, $p < .001$. The model explained 4.5% (Nagelkerke R2) of the variance in sentiment and correctly classified 73.1% of cases. Energy, Financial Services, and Industrial sectors were found to be between 2.1 and 2.5 times more likely to obtain a negative reaction than other sectors, with symbolic action 1.4 times and corporate praising 1.5 times.

The following table helps to understand how much variation in the dependent variable can be explained by the model. Based on the model, the explained variation in the dependent variable ranges from 3.1% to 4.5%, using the Cox & Snell R^2 and the Nagelkerke R^2, respectively.

Table 16 - Logistic regression model summary

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	6987.568[a]	.031	.045

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

The following table allows to assess the effectiveness of the predicted classification against the actual classification. Logistic regression estimates the probability of an event (in this case, having a negative reaction) occurring. If the estimated probability of the event is greater than or equal to 0.5, the event is classified as occurring (negative reaction). If the probability is less than 0.5, the event is classified as not occurring (positive reaction). The model has an overall accuracy of 73.1%, with 98.9% sensitivity and 5.0% specificity, meaning that the model accurately predicted if a corporate tweet would generate a negative response but not if a tweet would generate a positive response.

Table 17 - Logistic regression classification table

Classification Table[a]					
Observed			Predicted		
			prediction		Percentage Correct
			0	1	
Step 1	prediction	0	84	1594	5.0
		1	49	4381	98.9
	Overall Percentage				73.1

a. The cut value is .500

Table 18 - Logistic regression classification metrics

Accuracy	.0731
Precision	.733
Sensitivity	.989
Specificity	.050

The following table shows the contribution of each independent variable to the model and its statistical significance. The statistical significance of the test is found in the “Sig.” column. From these results, we can see that sector ($p < .001$) and action ($p < .001$) added significantly to the model.

Table 19 - Logistic regression variables in the equation

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	cdp_climate_change_score_2021_numeric	.022	.014	2.483	1	.115	1.023	.995	1.052
	net_zero_flag	-.126	.083	2.313	1	.128	.881	.749	1.037
	sector			152.986	6	<.001			
	sector(1)	.080	.360	.050	1	.823	1.084	.535	2.194
	sector(2)	.770	.359	4.617	1	.032	2.161	1.070	4.363
	sector(3)	.751	.399	3.552	1	.059	2.120	.970	4.630
	sector(4)	-.544	.702	.600	1	.439	.580	.146	2.299
	sector(5)	.917	.353	6.738	1	.009	2.502	1.252	5.002
	sector(6)	-.939	.402	5.465	1	.019	.391	.178	.859
	claim			4.234	2	.120			
	claim(1)	.136	.070	3.849	1	.050	1.146	1.000	1.314
	claim(2)	.131	.094	1.934	1	.164	1.139	.948	1.370
	action(1)	.338	.077	19.143	1	<.001	1.402	1.205	1.632
	praising(1)	.441	.257	2.940	1	.086	1.555	.939	2.575
	Constant	-.397	.448	.785	1	.376	.672		

a. Variable(s) entered on step 1: cdp_climate_change_score_2021_numeric, net_zero_flag, sector, claim, action, praising.

Lastly, the chi-square test was applied to observe the relation between sentiment and the other variables. The chi-square test assumes that the expected value for each cell is five or higher. Fisher's exact test is applied if this assumption is not met. The results indicate that there is a statistically significant relationship between sentiment and cdp climate change score ($\chi 2(6) = 61.166$, $p <.001$), sector (FET = 150.780, $p < .001$), claim ($\chi 2(2) = 9.326$, $p = .010$), action ($\chi 2(1) = 20.975$, $p < .001$) and praising ($\chi 2(1) = 5.161$, $p = .018$). The results also indicate no statistically significant relationship between sentiment and net zero flag ($\chi 2(1) = .017$, $p = .463$).

4.3. Strategies for effective marketing campaigns

This research question aims to elaborate on effective marketing campaigns, outlining the recommendations for campaigns that are unlikely to be perceived as greenwashing, identifying common threads among campaigns that have been met with little or no opposition, and identifying the factors that contribute to the opposition. Little to no opposition in this context means corporate tweets that have received mainly positive replies and joyful emotions. Overall, positive sentiment and joyful emotions represent 26% of the data.

The type of compensation claim has little to no effect on sentiment and emotion. It was found that abstract compensation claims generate 30% positive responses, concrete compensation claims generate 25%, and vague claims 28%. During the literature review, compensation claim was the only factor where authors diverged in their findings. Some assert that abstract, vague, and false claims contribute to higher levels of perceived greenwashing (Neureiter & Matthes, 2022), while others argue that vague claims in green advertising are unrelated to perceived greenwashing and do not enhance consumers skepticism compared to non-deceptive claims (Schmuck et al., 2018). The present research also finds compensation claims inconclusive and worthy of further studies.

Similarly to claims, actions have little or no effect on the consumer response (sentiment and emotions). Substantive actions generate 29% positive responses, while symbolic generate 25%. In the literature review, Zhang and Sun, 2021) noted that consumers generally respond more positively to substantive environmental actions than symbolic ones. Although the effect is negligible, this research also indicates that substantive actions are more effective in generating positive responses.

The most significant differences were in sentiment and emotions in the company sectors. Sectors with a very expressive positive reaction include basic materials, Consumer Cyclical and Consumer Defensive, Health Care, and Technology. In particular, Basic Materials and Health Care have an incredibly joyful response. This is merely an observation of the data. Although it might not help a company in a sector like Energy, it might help to manage expectations and go the extra mile to swift consumer perception.

Praising also shows significant differences. Consumer praising generates 39.3% positive reactions, while corporate praising only generates 27.7%. This difference is also evident in emotions, where consumer praising generates 44% joyful emotions while corporate praising generates 32%. Researchers (Christis and Wang, 2021) highlight that participants exposed to consumer praise displayed a higher tendency to trust the brand.

The presence of net zero or related terms did not have a statistically significant effect on consumer reactions.

On the other hand, the climate change score had an unexpected effect on consumer response. Positive sentiment tends to decrease as the climate change score increases. However, joyful emotions tend to increase slightly with the score.

To provide clear and meaningful strategies for effective marketing campaigns, some aspects of this research must be left aside for two main reasons: (1) this research found that they are not meaningful enough to be considered, and (2) to avoid being overly specific no recommendations will be made regarding each sector approach to develop effective marketing strategies.

This research has proven empirically that substantive actions generate more positive responses than symbolic ones, agreeing with the finding (Zhang and Sun, 2021). The companies in the dataset seem to be on the right path, with 65.63% of corporate tweets announcing substantive action on their part. In this research, substantive action was defined as concrete and visible actions in management goals, company structures, and institutionalization initiatives, where companies discuss their environmental responsibility in terms of what they are doing now or what they have done. Symbolic actions, on the other hand, were defined as an array of superficial, negligible, and easy-to-be-observed environmental gestures aiming to obtain external validation and social support, such as establishing an environmental supervision committee or using green labels or trademarks.

This research has also empirically proven that consumer praising generates more positive reactions than corporate praising. Although assessing consumer trust towards the brand was out of the scope of this research, it is safe to agree with (Christis and Wang, 2021) that consumer praise generates trust in consumers. However, unlike substantive actions, consumer praising is rarely seen in this dataset, with only 3.55% of corporate tweets demonstrating this praise. In this research, corporate praising was defined as praising the company for its social actions, such as its sustainability initiatives. In contrast, consumer praising was defined as praising consumers for their sustainable behavior or purchases.

CHAPTER 5

Conclusions

5.1. Conclusion

This research aimed to answer three research questions: 1) the extent to which consumer responses to online campaigns are focused on greenwashing or other issues, 2) determine the feelings consumers demonstrate towards such campaigns, and determine if feelings are dependent on a variety of observable factors and 3) to outline recommendations for effective marketing campaigns that are unlikely to be perceived as greenwashing. The analysis was conducted using the CRISP-DM methodology, and data was collected from various sources, including Twitter, Morning Star, and the CDP website.

The literature review process allowed for a deeper understanding of the concepts involved in greenwashing, consumer behavior, and environmental communication. Based on inclusion and exclusion criteria, 23 articles were selected out of 125. Most research in green marketing and greenwashing has been empirical and quantitative, with a large proportion based on survey methodology. The literature review also revealed concepts that would be later used to enhance this research, such as type of compensation claim, praising, and action. Thus, this research aimed to fill a gap in the literature by conducting a broad analysis (without a specific focus on a particular company, event, or industry) of potential consumers' reactions to greenwashing. It intended to add a new dimension to previously conducted studies by using NLP (natural language processing) techniques to differentiate between positive and negative reactions and identify emotions such as sadness, joy, or disgust in said reactions. The body of research in this area is scarce, indicating significant room for further investigation and development. By analyzing consumer opinions and sentiments towards greenwashing, this research aimed to shed light on the issue and provide insights for companies to improve their marketing efforts and avoid the perception of greenwashing.

This research is primarily descriptive because it investigates the implications of customers' reactions to greenwashing practices.

Multiple methods and visualizations were used to answer the first research question (consumer response analysis). The first was Topic Modeling LDA, a generative unsupervised probabilistic algorithm used to discover hidden thematic patterns within a collection of documents. Applying this algorithm to the replies reveals three topics: one focused on climate issues, another on supporting causes and initiatives, and lastly, one about customer service. The second analysis was Document Frequency (DF), used to understand the importance of words within a collection of documents. This analysis revealed that the most important words were customer, people, flight, money, and service. Climate-related words were in the background. Next was N-gram analysis, used to capture patterns and relationships between words or characters in a text. Although customer service was pronounced this time, climate change and fossil fuel were also significantly used in replies. Lastly, a dictionary analysis was conducted. This revealed that only 13.9% of replies refer to climate issues, most of those in the Energy and Financial Services sectors and those with a CDP score of zero for climate change.

To answer the second research question (consumer sentiment analysis), sentiment analysis and emotion detection techniques were applied. This research adopted the transfer learning approach (using pre-trained models in other similar domains). Specifically, it used the BERT language model to distinguish between positive, negative, and neutral sentiments and six different emotions (anger, disgust, fear, joy, sadness, and surprise). Most replies had a negative sentiment, although joy was also a prominent emotion. The literature reveals that greenwashing can negatively impact consumer emotions as perceived greenwashing has been linked to decreased happiness (Szabo & Webster, 2021). The research found that the type of compensation claim, similar to the presence of net zero terms, had little or no effect on sentiment and emotion. The literature review was inconclusive on the inconclusive of compensation claim, Neureiter & Matthes (2022) found that abstract, vague, and false claims contribute to higher levels of perceived greenwashing, while Schmuck et al. (2018) found that vague claims in green advertising are unrelated to perceived greenwashing and do not enhance consumers skepticism compared to non-deceptive claims. Although its effect is negligible, the type of action can be considered a factor that shapes consumer sentiment, with substantive action generating a more positive response. This finding aligns with Zhang and Sun (2021) that consumers generally respond more positively to companies' substantive environmental actions than symbolic ones. Conversely, the sector has a significant impact, with Communication Services, Energy, Financial Services, and Industrial generating a significant negative response. Praising also has a meaningful impact, with corporate praising generating a much stronger negative response when compared to consumer praising. This finding aligns with Christis and Wang (2021), highlighting that participants exposed to consumer praise tended to trust the brand more. Lastly, the climate change score had an unexpected effect, indicating that an increase in climate score is associated with a slight increase in negative sentiment. However, the most significant differences were found in emotions, where anger decreased significantly as climate score increased while sadness and joy increased. Two further methods were used to confirm the statistical significance of the above analysis: Logistic Regression and Chi-square test. The logistic regression model was statistically significant ($\chi2(12) = 194.203$, $p < .001$), explaining 4.5% (Nagelkerke R2) of the variance in sentiment and correctly classified 73.1% of cases. This analysis proved that the Energy, Financial Services, and Industrial sector were between 2.1 and 2.5 times more likely to obtain a negative reaction than other sectors. Similarly, symbolic action was 1.4 times, and corporate praising was 1.5 times more likely to obtain a negative reaction. Lastly, the chi-square test confirmed a significant association between sentiment and the remaining variables discussed previously, apart from the net zero flag.

The final research question (strategies for effective marketing campaigns) elaborated on these research findings to create an actionable plan to elaborate marketing campaigns. It was concluded that substantive action and consumer praising are more viable approaches to generate positive reactions and thus minimize greenwashing.

5.2. Limitations

One possible limitation of this research is the CDP climate score. As pointed out by (Johnson & Greenwell, 2022), it is unclear whether the score is entirely fair and what a fair score would look like. For example, one huge energy company was granted a high score in 2019 (B: Climate management) despite being one of the ten largest carbon-emitting companies and releasing nearly 2% of the world's greenhouse gasses. They raise essential dilemmas, such as should these extreme emitters be granted high scores? Should companies be rewarded with a high score for mitigating carbon emissions, even if their mitigation had a negligible impact when considering their overall emissions?

Another possible limitation is the limited representation of climate-related concerns in negative sentiment. While the sentiment analysis of consumer replies to corporate tweets identified as greenwashing revealed a predominant negative sentiment, it is essential to acknowledge the limited representation of climate-related concerns within these negative responses. Despite the overarching negative sentiment observed, only 13.9% of the replies explicitly mention climate issues or concerns related to environmental claims made by the companies. This limitation highlights the complexity of consumer attitudes and the diverse range of reasons that contribute to negative responses. Consumers' negative sentiment may stem from various factors beyond climate-related concerns. Consequently, the research findings may not fully capture the nuances of consumer skepticism toward greenwashing practices.

A subsequent limitation is the implications of the limited representation of climate concerns. An inherent limitation of this research lies in the relatively low representation of climate-related concerns within the negative sentiment expressed by consumers in response to corporate tweets identified as greenwashing. Despite the notable prevalence of negative sentiment, only a minority of the replies (13.9%) explicitly mentioned climate issues or environmental concerns. This discrepancy between the overall negative sentiment and the specific focus on climate-related topics raises questions about how the recommendations proposed for mitigating negative reactions directly apply to the broader greenwashing problem. The dearth of explicit climate-related mentions in consumer responses could be attributed to several factors, including the complexity of consumer sentiment, varying consumer priorities, and the limited scope of Twitter's character constraints for expressing multifaceted concerns. As a result, the recommendations, such as emphasizing substantive actions and fostering consumer praise, may appear tailored to a subset of consumer concerns rather than comprehensively addressing the broader issue of greenwashing. Furthermore, the absence of extensive climate-related discourse in consumer replies poses challenges in determining the true impact of corporate greenwashing on shaping consumer sentiment. While the negative sentiment expressed may stem from diverse reasons, the research does not have a comprehensive view of whether consumers know the underlying greenwashing attempts or whether their responses reflect broader skepticism towards corporate practices. To mitigate this limitation, future research endeavors could include more diverse data sources beyond Twitter and employ qualitative methodologies to delve deeper into consumer perceptions and motivations behind negative sentiment. This could provide a more nuanced understanding of the interplay between greenwashing and consumer reactions, shedding light on the underlying factors influencing sentiment and guiding their responses. Moreover, a longitudinal study could track changes in consumer sentiment over time, revealing whether climate concerns gain prominence as environmental awareness continues to evolve. In conclusion, while the current research recommendations for a refined marketing strategy offer valuable insights, it is essential to acknowledge the limited representation of climate concerns in the analyzed consumer replies. This limitation underscores the need for continued research to explore the multifaceted dimensions of consumer attitudes toward greenwashing and to refine strategies that authentically resonate with an evolving range of consumer concerns.

Bibliography

Baum, L. M. (2012). It's not easy being green...or is it? a content analysis of environmental claims in magazine advertisements from the United States and United Kingdom. *Environmental Communication*, *6*(4), 423–440. https://doi.org/10.1080/17524032.2012.724022

Beriain Bañares, A., Santos Silva, M. F., & Rodríguez Rodríguez, S. (2021). Green but ignored? The irrelevance of television advertisements on energy sustainability in Spain and its impact on consumer perceptions. *Energy Research and Social Science*, *73*. https://doi.org/10.1016/j.erss.2020.101835

Berthold, M. R., Borgelt, C., Höppner, F., Klawonn, F., & Silipo, R. (2020). *Guide to Intelligent Data Science*. Springer International Publishing. https://doi.org/10.1007/978-3-030-45574-3

Christis, J., & Wang, Y. (2021). Communicating environmental csr towards consumers: The impact of message content, message style and praise tactics. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073981

de Alencar Caldas, M. V., Veiga-Neto, A. R., de Almeida Guimarães, L. G., de Castro, A. B. C., & Pereira, G. R. B. (2021). Greenwashing in environmental marketing strategy in the brazilian furniture market. *Revista de Economia e Sociologia Rural*, *59*(3), 01–19. https://doi.org/10.1590/1806-9479.2021.225336

De Jong, M. D. T., Harkink, K. M., & Barth, S. (2018). Making Green Stuff? Effects of Corporate Greenwashing on Consumers. *Journal of Business and Technical Communication*, *32*(1), 77–112. https://doi.org/10.1177/1050651917729863

de Jong, M. D. T., Huluba, G., & Beldad, A. D. (2020). Different Shades of Greenwashing: Consumers' Reactions to Environmental Lies, Half-Lies, and Organizations Taking Credit for Following Legal Obligations. *Journal of Business and Technical Communication*, *34*(1), 38–76. https://doi.org/10.1177/1050651919874105

Delmas, M. A., & Burbano, V. C. (n.d.). *The Drivers of Greenwashing*.

Fankhauser, S., Smith, S. M., Allen, M., Axelsson, K., Hale, T., Hepburn, C., Kendall, J. M., Khosla, R., Lezaun, J., Mitchell-Larson, E., Obersteiner, M., Rajamani, L., Rickaby, R., Seddon, N., & Wetzer, T. (2022). The meaning of net zero and how to get it right. In *Nature Climate Change* (Vol. 12, Issue 1, pp. 15–21). Nature Research. https://doi.org/10.1038/s41558-021-01245-w

Fernandes, J., Segev, S., & Leopold, J. K. (2020). When consumers learn to spot deception in advertising: testing a literacy intervention to combat greenwashing. *International Journal of Advertising*, *39*(7), 1115–1149. https://doi.org/10.1080/02650487.2020.1765656

Guerreiro, J., & Pacheco, M. (2021). How green trust, consumer brand engagement and green word-of-mouth mediate purchasing intentions. *Sustainability (Switzerland)*, *13*(14). https://doi.org/10.3390/su13147877

Hartmann, P., & Apaolaza-Ibáñez, V. (2012). Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern. *Journal of Business Research*, *65*(9), 1254–1263. https://doi.org/10.1016/j.jbusres.2011.11.001

Johnson, T. F., & Greenwell, M. P. (2022). Are companies using Twitter to greenwash and hide bad environmental performance? *Energy, Ecology and Environment*, *7*(3), 213–226. https://doi.org/10.1007/s40974-021-00237-y

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.

Knight, H., Haddoud, M. Y., & Megicks, P. (2022). Determinants of corporate sustainability message sharing on social media: A configuration approach. *Business Strategy and the Environment*, *31*(2), 633–647. https://doi.org/10.1002/bse.2941

Lu, X., Sheng, T., Zhou, X., Shen, C., & Fang, B. (2022). How Does Young Consumers' Greenwashing Perception Impact Their Green Purchase Intention in the Fast Fashion Industry? An Analysis from the Perspective of Perceived Risk Theory. *Sustainability (Switzerland)*, *14*(20). https://doi.org/10.3390/su142013473

Lyon, T. P., & Montgomery, A. W. (2013). Tweetjacked: The Impact of Social Media on Corporate Greenwash. *Journal of Business Ethics*, *118*(4), 747–757. https://doi.org/10.1007/s10551-013-1958-x

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Munaier, C. G. e. S., Miyazaki, F. R., & Mazzon, J. A. (2022). Morally transgressive companies and sustainable guidelines: seeking redemption or abusing trust? *RAUSP Management Journal*, *57*(4), 413–433. https://doi.org/10.1108/RAUSP-01-2022-0047

Musgrove, C. (Casey) F., Choi, P., & Chris Cox, K. (2018). Consumer Perceptions of Green Marketing Claims: An Examination of the Relationships with Type of Claim and Corporate Credibility. *Services Marketing Quarterly*, *39*(4), 277–292. https://doi.org/10.1080/15332969.2018.1514794

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. In *Social Network Analysis and Mining* (Vol. 11, Issue 1). Springer. https://doi.org/10.1007/s13278-021-00776-6

Neureiter, A., & Matthes, J. (2022). Comparing the effects of greenwashing claims in environmental airline advertising: perceived greenwashing, brand evaluation, and flight shame. *International Journal of Advertising*. https://doi.org/10.1080/02650487.2022.2076510

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, *27*(7), 1326–1334. https://doi.org/10.1080/02699931.2013.777660

Oliveira, V. M. De, Aguiar, E. C., Melo, L. S. A., & Correia, S. É. N. (2019). Marketing and Green Consumption: Greenwashing Influence on Green Consumer Trust. *Revista de Gestao Social e Ambiental*, *13*(2), 93–110. https://doi.org/10.24857/RGSA.V13I2.2038

Parguel, B., Benoit-Moreau, F., & Russell, C. A. (2015). Can evoking nature in advertising mislead consumers? The power of 'executional greenwashing.' *International Journal of Advertising*, *34*(1), 107–134. https://doi.org/10.1080/02650487.2014.996116

Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, R., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.

Quinn, J. (2020). *PREDICTIVE ANALYTICS Jarlath Quinn Smart Vision Europe*.

Reck, R., Castagna, A. C., Shuqair, S., & Costa Pinto, D. (2022). The transparency paradox: When transparency cues helps or backfires for brands? *Journal of Cleaner Production*, *372*. https://doi.org/10.1016/j.jclepro.2022.133381

Sailer, A., Wilfing, H., & Straus, E. (2022). Greenwashing and Bluewashing in Black Friday-Related Sustainable Fashion Marketing on Instagram. *Sustainability (Switzerland)*, *14*(3). https://doi.org/10.3390/su14031494

Schmuck, D., Matthes, J., & Naderer, B. (2018). Misleading Consumers with Green Advertising? An Affect–Reason–Involvement Account of Greenwashing Effects in Environmental Advertising. *Journal of Advertising*, *47*(2), 127–145. https://doi.org/10.1080/00913367.2018.1452652

ShabbirHusain, R. V., & Varshney, S. (2019). Is Current Way of Promoting Sustainability, Sustainable? *Journal of Nonprofit and Public Sector Marketing*, *31*(1), 84–113. https://doi.org/10.1080/10495142.2018.1526735

Singh, N., Gupta, K., & Kapur, B. (2022). Greenwashed word of mouth (GWWOM): a vibrant influence on customer green behaviour. *Journal of Global Responsibility*, *13*(4), 472–487. https://doi.org/10.1108/JGR-11-2021-0094

Sun, Y., & Shi, B. (2022). Impact of Greenwashing Perception on Consumers' Green Purchasing Intentions: A Moderated Mediation Model. *Sustainability (Switzerland)*, *14*(19). https://doi.org/10.3390/su141912119

Szabo, S., & Webster, J. (2021). Perceived Greenwashing: The Effects of Green Marketing on Environmental and Product Perceptions. *Journal of Business Ethics*, *171*(4), 719–739. https://doi.org/10.1007/s10551-020-04461-0

Tursi, V., & Silipo, R. (2021). *KNIME PRESS TEXT MINING*.

Walker, K., & Wan, F. (2012). The Harm of Symbolic Actions and Green-Washing: Corporate Actions and Communications on Environmental Performance and Their Financial Implications. *Journal of Business Ethics*, *109*(2), 227–242. https://doi.org/10.1007/s10551-011-1122-4

Zhang, H., Ul Ainn, Q., Bashir, I., Ul Haq, J., & Bonn, M. A. (2022). Does Greenwashing Influence the Green Product Experience in Emerging Hospitality Markets Post-COVID-19? *Sustainability (Switzerland)*, *14*(19). https://doi.org/10.3390/su141912313

Zhang, J., & Sun, J. (2021). Green talk or green walk: Chinese consumer positive word-of-mouth to corporate environmental actions in polluting industries. *Sustainability (Switzerland)*, *13*(9). https://doi.org/10.3390/su13095259

Zhang, L., Li, D., Cao, C., & Huang, S. (2018). The influence of greenwashing perception on green purchasing intentions: The mediating role of green word-of-mouth and moderating role of green concern. *Journal of Cleaner Production*, *187*, 740–750. https://doi.org/10.1016/j.jclepro.2018.03.201

Annexes

Annex A – Climate Dictionary

greenwash
greenwashing
climate
global warming
climate change
1.5 pathway
anthropocene
anthropogenic
carbon dioxide
co2
carbon emissions
carbon footprint
carbon sequestration
carbon stock
carbon tax
carbon
biodiversity
climate crisis
climate emergency
climate justice
climate overshoot
climate velocity
ecoanxiety
ecolinguistics
extreme weather events
megadrought
drought
megafire
fire
mitigation
mass extinction event
paris agreement
sea-level rise
greenhouse gases
emissions
tipping point
climate feedback loops
feedback loop
climate security
net zero
net-zero
decarbonization
renewable energy
carbon sink
carbon removal
carbon markets
regenerative agriculture
afforestation
rewilding
circular economy
blue economy
green jobs
cop
abrupt climate change
adaptation
adaptive capacity
aerosols
albedo
alternative energy
atmosphere
atmospheric lifetime
biofuels
biogeochemical cycle
biomass
biosphere
black carbon aerosol
borehole
carbon cycle
carbon dioxide
carbon dioxide equivalent
carbon dioxide fertilization
carbon capture and sequestration
chlorofluorocarbons
climate feedback
climate lag
climate model
climate sensitivity
climate system
coal mine methane
coalbed methane
co-benefit
concentration
conference of the parties
coral bleaching
cryosphere
deforestation
desertification
eccentricity
dryland farming
ecosystem
emissions factor
energy efficiency
greenhouse effect
fluorinated gases
fluorocarbons
fossil fuel
geosphere
glacier
global average temperature
global warming potential
greenhouse gas
heat waves
hydrocarbons
hydrochlorofluorocarbons
hydrofluorocarbons
hydrologic cycle
hydrosphere
indirect emissions
infrared radiation
landfill
methane
nitrogen oxides
nitrous oxide
non-methane volatile organic compounds
ocean acidification
oxidize
ozone
ozone depleting substance
ozone layer
particulate matter
perfluorocarbons
photosynthesis
recycling
reforestation
relative sea level rise
renewable energy
sea surface temperature
soil carbon
solar radiation
stratosphere
sulfate aerosols
sulfur hexafluoride
water vapor
weather
offset emissions
carbon offset
bluewashing
toxic
renewables

Annex B – Chi-square test

Chi-Square Tests (cdp_climate_change_score_2021_numeric)						
	Value	df	Asymptotic Significance (2-sided)	Monte Carlo Sig. (2-sided)		
				Significance	99% Confidence Interval	
					Lower Bound	Upper Bound
Pearson Chi-Square	61.166[a]	6	<.001	<.001[b]	<.001	<.001
Likelihood Ratio	57.274	6	<.001	<.001[b]	<.001	<.001
Fisher-Freeman-Halton Exact Test	57.590			<.001[b]	<.001	<.001
N of Valid Cases	6108					

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 13.74.

b. Based on 10000 sampled tables with starting seed 1502173562.

Chi-Square Tests (net_zero_flag)					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.017[a]	1	.896	.915	.463
Continuity Correction[b]	.009	1	.924		
Likelihood Ratio	.017	1	.896	.915	.463
Fisher's Exact Test				.915	.463
N of Valid Cases	6108				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 339.83.

b. Computed only for a 2x2 table

Chi-Square Tests (sector)						
	Value	df	Asymptotic Significance (2-sided)	Monte Carlo Sig. (2-sided)		
				Significance	99% Confidence Interval	
					Lower Bound	Upper Bound
Pearson Chi-Square	164.698[a]	7	<.001	<.001[b]	<.001	<.001
Likelihood Ratio	149.923	7	<.001	<.001[b]	<.001	<.001
Fisher-Freeman-Halton Exact Test	150.780			<.001[b]	<.001	<.001
N of Valid Cases	6108					

a. 1 cells (6.3%) have expected count less than 5. The minimum expected count is 3.02.

b. Based on 10000 sampled tables with starting seed 112562564.

Chi-Square Tests (claim)				
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	9.326[a]	2	.009	.010
Likelihood Ratio	9.354	2	.009	.009
Fisher-Freeman-Halton Exact Test	9.343			.009
N of Valid Cases	6108			

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 297.80.

Chi-Square Tests (action)					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	20.975[a]	1	<.001	<.001	<.001
Continuity Correction[b]	20.686	1	<.001		
Likelihood Ratio	21.417	1	<.001	<.001	<.001
Fisher's Exact Test				<.001	<.001
N of Valid Cases	6108				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 480.21.

b. Computed only for a 2x2 table

Chi-Square Tests (praising)					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.161[a]	1	.023	.026	.018
Continuity Correction[b]	4.583	1	.032		
Likelihood Ratio	4.808	1	.028	.035	.018
Fisher's Exact Test				.026	.018
N of Valid Cases	6108				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 20.33.

b. Computed only for a 2x2 table