Sentiment Analysis of New Zealand Adults' and Children's Tweets Regarding the COVID-19 Vaccination Programme

Lamyaa Aldahmash

2023

A dissertation submitted to Auckland University of Technology in partial fulfilment of the requirements for the degree of Master of Public Health

School of Public Health and Interdisciplinary Studies

Abstract

The SARS-CoV-2 virus, which caused the global COVID-19 pandemic, necessitated a significant worldwide response, with vaccination being a primary strategy. This dissertation explores the public sentiment towards New Zealand's national vaccination campaign, through a machine learning analysis of large-scale text data gathered from the social media platform Twitter. Focusing on responses from both adults and children, this research aimed to assess the efficacy of health communication strategies and the wider acceptance of the vaccine within the community. The findings underscore a considerable disparity between policy decisions and public sentiment on Twitter, with a significant portion of the New Zealand population expressing negative views on vaccinations. Overall, this research reveals the need for enhanced public engagement, better communication, and more effective use of social media data by policymakers and healthcare professionals in order to address public concerns, mitigate fears, dispel misinformation, and ultimately increase vaccine uptake.

Table of Contents

List of Abbreviations

API	Application Programming Interfaces
BERT	Bidirectional Encoder Representations from Transformers
CEO	Chief Executive Officer
EDA	Exploratory Data Analysis
HPV	Human Papillomavirus Vaccines
MBIE	Ministry of Business, Innovation, and Employment
MERS	Middle East Respiratory Syndrome
MMR	Measles, Mumps and Rubella
NLP	Natural Language Processing
NZPSU	New Zealand Pediatric Surveillance Unit
OECD	Organisation for Economic Cooperation and Development
SARS	Severe Acute Respiratory Syndrome
WHO	World Health Organization

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature

Date 1st July 2023

Acknowledgements

The completion of this research would not have been possible without the gracious assistance of many individuals.

First and foremost, I would like to express my deepest gratitude to my primary supervisor, Charles Mpofu, whose unwavering support and guidance have been invaluable throughout all stages of this research and the writing of my dissertation. His direction, inspiration, persistence and extensive knowledge have played a vital role in enabling me to complete this dissertation.

I am particularly indebted to Dr Abdulaziz Alaboudi for his generous support and guidance in working with the software used in this research. Our collaborative effort has been a fruitful exchange of knowledge and ideas, and I am truly grateful for his contributions.

I would also like to acknowledge my husband for his unending dedication to the welfare of our children and his willingness to stand by my side throughout the challenges I faced in the pursuit of knowledge.

To my family, I owe an immeasurable debt of gratitude. They have provided me with the confidence, strength, encouragement and courage to pursue this goal. I want to specifically thank my dear mother, brothers and sisters for instilling in me the value of education and for being a constant source of inspiration.

I also extend my heartfelt appreciation to all my friends and individuals from different parts of the country who have supported me throughout my academic journey. Their unwavering support has been a source of motivation and encouragement.

Lastly, I am truly grateful for the assistance, guidance and support of all those who have played significant roles in my academic and personal life. Their contributions have been invaluable, and I am humbled by their kindness and generosity.

Chapter 1. Introduction

1.1 Overview

On 11 March 2020, the World Health Organization (WHO) acknowledged the severity and spread of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, classifying the disease it causes, COVID-19, as a global pandemic. The virus swiftly infiltrated nearly every corner of the world, prompting a comprehensive and urgent global response (WHO, 2020). Insightful research conducted by the US National Library of Medicine (2022) reveals that, in response to this health crisis, nations around the world devised a myriad of strategies to curtail the pervasive virus. The primary weapon in this battle has been vaccination, a biotechnological marvel that has been globally recognised and applied (National Library of Medicine, 2022).

New Zealand serves as a notable case study in the global effort against the virus. The country, recognised for its prompt and effective response to the pandemic, launched an intensive national vaccination campaign in 2021. This program, backed by rigorous scientific research, aimed to accelerate the vaccination process across all age groups. This comprehensive approach was undertaken with the goal of securing herd immunity and thereby, ensuring the safety and wellbeing of the entire nation (New Zealand Ministry of Health, 2022).

But how well was New Zealand's vaccination campaign received by its people? The research presented in this dissertation focused on the sentiments of the New Zealand population towards the vaccination program, capturing perspectives from both adults and children. An understanding of public sentiment towards vaccination is crucial in gauging the efficacy of health communication strategies and the broader acceptance of vaccines within the community (Betsch et al., 2018; Brewer et al., 2017).

The research employed a machine learning approach, an innovative and increasingly valued methodology in the field of social science research. Machine learning algorithms can be leveraged to analyse large-scale text data from various digital platforms, pinpointing trends and patterns in public sentiment that would otherwise be challenging to discern (Grimmer & Stewart, 2013; Nelson, 2020). This methodology offers the possibility to dissect public opinion in a nuanced, efficient, and comprehensive way, accommodating the complexity and diversity of opinions that exist within a population (Bail et al., 2020).

This research contributes to the growing body of literature that employs machine learning techniques in social science research with a specific focus on public health communication in the context of a global pandemic.

1.2 COVID-19 Pandemic and Public Perception

Dr Seth Berkley, CEO of the Geneva-based Gavi, the Vaccine Alliance, articulated a poignant message amidst the tumultuous backdrop of the COVID-19 pandemic: “We cannot have an equitable recovery from the pandemic if we leave the most vulnerable in our communities behind” (WHO, 2022). This remark amplifies the importance of investigating and understanding the public sentiment surrounding vaccination policies, especially among the most vulnerable our children. The global turbulence wrought by the pandemic necessitates strategies that researchers consider the entire demographic spectrum, including adults and children. A comprehensive understanding of public sentiment can aid in formulating and implementing policies that address the needs and concerns of all, ensuring no group is inadvertently overlooked.

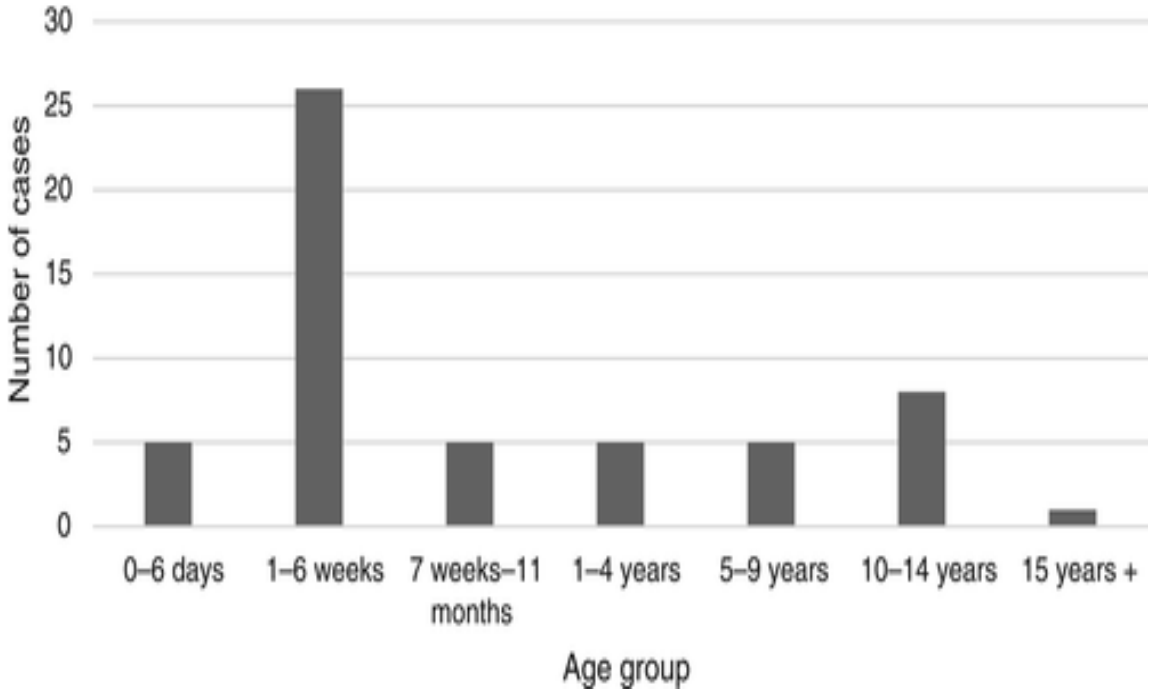
In the context of New Zealand, the need for such sentiment analysis is pronounced. A study conducted by the New Zealand Paediatric Surveillance Unit (NZPSU) sheds light on the impact of the pandemic on children (Duncanson et al., 2021). Of the reported cases of compromised care, a significant number pertained to infants and young children, emphasising the urgency of their inclusion in the national recovery plan (Duncanson et al., 2021).

Figure 1 shows the spread of COVID-19 cases among different age groups of children in New Zealand (Duncanson et al., 2021). The researchers found infants in the first year of life accounted for 65% (36 cases) of their sample. The majority of these cases were infants under 6 weeks old. In terms of older age groups, the cases were evenly distributed (Duncanson et al., 2021).

This finding is a stark reminder that the disruption caused by COVID-19 extends far beyond health, affecting the wellbeing of children, and highlights the urgent need for vaccination strategies attuned to the needs and sentiments of this demographic.

Figure 1

Distribution of COVID-19 cases across different age groups of New Zealand children



Source: Duncanson et al. (2021).

The central objective of this research is therefore to use machine learning approaches to analyse sentiment towards COVID-19 vaccination policies within the adult and child populations in New Zealand. This research aims to generate valuable insights that can inform equitable and effective health strategies in the fight against future pandemics.

1.3 Motivations, Rationale, Aims, Research Questions and Constraints of the Research

1.3.1 Motivations for Selecting This Topic

The motivation for this research arose from a desire to help fill the existing void of knowledge about the pandemic's repercussions on individuals' lives and wellbeing, particularly within New Zealand. While substantial efforts have been directed globally towards developing strategies to

curb the pandemic, the effects of these measures on the day-to-day lives, mental health and overall wellness of individuals have remained largely under-examined (Duden, 2022).

In the context of New Zealand, this subject takes on additional weight due to the country's distinctive socio-cultural environment and its unique public health responses to the pandemic (Henrickson, 2020). For instance, New Zealand's swift lockdown measures and strict quarantine rules, while successful in controlling the spread of the virus, may have had unforeseen impacts on the population's mental health, economic circumstances, and social dynamics.

This researcher is particularly intrigued by these potential impacts and how they relate to overall wellbeing. The desire to illuminate this lesser-explored domain, to comprehend the repercussions of policy decisions on the population's wellbeing, and to suggest possible enhancements, served as the key motivation for the selection of this topic. In-depth research in this field may ultimately guide policymakers in making more holistic decisions that consider not only the direct health impacts of a pandemic, but also its broader social, economic and psychological effects.

1.3.2 Research Rationale

A key inspiration for this research was Jacinda Ardern, the former prime minister of New Zealand. In the face of a rapidly developing global pandemic, she appealed for "cooperation, communication, and compassion" (May, 2021). The shifts in New Zealand's approach from elimination to mitigation to "learning to live" with the virus should inform any investigation designed to evaluate policy effectiveness. Many vulnerable groups of the population may have been left to bear the danger (Baker & Wilson, 2022). As a responsible global citizen, it was natural to dwell on the investigation to find if the sentiments of the adults and children were positive or negative.

1.3.3 Research Aims

This research has four aims:

1. To gauge public sentiment towards New Zealand's vaccination campaign.
2. To utilise machine learning methodologies for large-scale text data analysis, drawing out patterns and trends in public sentiment.

3. To contribute to the literature on the use of machine learning techniques in social sciences, with a specific focus on public health communication.
4. To inform future health communication strategies based on the findings from the sentiment analysis.

1.3.4 Constraints

1. The scope of this research is limited to the region of New Zealand.
2. The targeted audience for the research purpose is adults and children within New Zealand.
3. The targeted audience for research purposes will be restricted to the social media platform Twitter.

1.3.5 Research Questions

After conducting an extensive literature review (see Chapter 2) to determine gaps in the knowledge within the scope of this study, the following two research questions were formulated:

1. How does public sentiment in New Zealand towards the COVID-19 vaccine differ between adult and paediatric vaccines as indicated by sentiment analysis?
2. To what extent are sentiments expressed on Twitter by adults and individuals presumed to be under the age of 18 related to the COVID-19 pandemic in New Zealand similar or different?

1.4 Challenges in Data Representation and Cleaning in Machine Learning Applications

Integrating machine learning with social science research, specifically the analysing of public sentiment towards COVID-19 vaccinations in New Zealand, presented distinct challenges. These predominantly related to data cleaning and ensuring representative data integrity.

1.4.1 Ensuring Data Representativeness

This study hinged on the analysis of sentiments expressed on social media, necessitating a dataset that authentically reflected New Zealand's diverse public opinion. A critical challenge was

overcoming the representation bias inherent in social media data. Not all segments of the population engage on these platforms, which could skew the sentiment analysis results. To address this, the study employed a systematic approach, incorporating demographic information and syntax variables to weigh and adjust the data. This rigorous refinement aimed to ensure the dataset was indicative of wider societal sentiment.

1.4.2 Rigorous Data Cleansing

Social media data is notoriously cluttered and noisy. Research of this kind involves the cleaning vast amounts of raw data to extract usable insights. This process involved filtering out irrelevant content, such as web links (i.e., URLs) and hashtags, crucial for accurate sentiment analysis. The machine learning algorithms used were tailored to distinguish nuances specific to sentiments on vaccinations, a task complicated by the informal communication style prevalent on social media.

1.4.3 Balancing Quantitative Output with Qualitative Richness

The final challenge was maintaining the depth and richness of a qualitative social science inquiry while leveraging the quantitative advantages of machine learning. This necessitated a balanced approach that could numerically evaluate sentiment trends without losing the individual narrative threads that qualitative analysis traditionally captures. To this end, the research methodology was designed to contextualise the large-scale data insights within the broader societal and cultural ethos of New Zealand during the pandemic.

By systematically addressing these challenges, the study ensured a robust analytical foundation, allowing for a more accurate, representative, and nuanced understanding of public sentiment towards COVID-19 vaccinations in New Zealand.

1.5 Dissertation Structure and Organisation

After this introductory chapter, the rest of this dissertation is structured into five chapters, which are summarised below:

Chapter 2. Literature Review: This chapter reviews the relevant literature in relation to the scope of this research in order to develop the research questions.

Chapter 3. Methodology: This chapter provides the scientific basis for conducting research to answer the research questions. It explains what data is to be collected, the methods used to collect the data, and justifies the chosen methodology.

Chapter 4. Findings: This chapter presents the findings of the sentiment analysis of tweets by adults and young people in New Zealand during the COVID-19 vaccination campaign.

Chapter 5. Discussion: This chapter discusses the results of this research in relation to the findings of the literature review. The limitations of the research are considered and a final decision will be made about the outcome of the study.

Chapter 6. Conclusion: This chapter summarises the conclusions of the study, discusses their implications and makes recommendations for future research.

1.6 Summary

This introductory chapter has provided an overview of the research conducted for this dissertation and the background to this study. The motivations, rationale, aims, research questions and constraints of the research were presented, along with an outline of the structure and organisation of the dissertation.

Chapter 2. Literature Review

2.1 Introduction

This chapter surveys the wealth of existing literature related to the public sentiments associated with the COVID-19 pandemic, with a particular emphasis on both global and New Zealand-specific perspectives. The study of public sentiment is crucial in understanding the multifaceted responses to this unprecedented global health crisis, shedding light on societal attitudes towards measures such as vaccination campaigns, lockdowns, and other public health strategies.

The next section, 'General Sentiments Relating to the COVID-19 Pandemic around the World', provides a broad overview of global public sentiment towards the pandemic. This section explores studies carried out across various nations with the aim of gaining a holistic perspective on global reactions to the pandemic and the strategies employed to combat it. Following this, the review narrows its focus in Section 2.3 to 'Adult Sentiments on Twitter Relating to the Pandemic in New Zealand and the World'. This section examines how adults, both within New Zealand and globally, have expressed their thoughts and feelings about the pandemic on Twitter. By investigating these discussions, this section offers insights into the broad range of opinions held by adults, a demographic that forms a significant portion of the Twitter-using population.

Section 2.4 then explores 'The Impact of the Pandemic on Young People in New Zealand and the World'. This section elucidates the specific challenges and sentiments expressed by younger demographics in the face of the pandemic. This focus is crucial, as young people's experiences and reactions may significantly differ from those of adults due to their unique socio-cultural and developmental contexts. In collating and examining these three distinct yet interconnected facets of pandemic-related sentiment, this literature review aims to provide a comprehensive and nuanced understanding of public responses to the COVID-19 pandemic.

2.2 Overview of Methodological Approaches in the Literature

In the pursuit of understanding public sentiment surrounding the COVID-19 pandemic, various methodological approaches have been employed by researchers. These methodologies can

be broadly categorised into machine learning-based techniques, sentiment analysis, and surveys. This section explores each of these methodologies, highlighting their strengths, limitations, and their degree of alignment with the research approach adopted in this study.

2.2.1 Machine Learning Approaches

Machine learning (ML) has emerged as a potent tool in the analysis of large datasets, such as social media data due to is ability to extract meaningful insights regarding public sentiment. Several studies reviewed for this research employed machine learning techniques to analyse Twitter data, among other platforms. These techniques include supervised learning, where models are trained on labelled data to classify sentiments, and unsupervised learning, where patterns are discerned from unlabelled data.

The strengths of machine learning include its efficiency in handling vast datasets; its ability to uncover complex patterns and trends; and its potential for real-time analysis. Its limitations include the requirement of substantial labelled data for supervised learning; the possible biases inherited from training data; and a dependence on data quality and representativeness.

2.2.2 Sentiment Analysis Techniques

Sentiment analysis (SA) encompasses a range of techniques aimed at extracting, identifying, or characterising the sentiment content of textual data. Techniques such as natural language processing (NLP) and lexicon-based approaches are prevalent in the literature.

The strengths of sentiment analysis include its ability to automate sentiment analysis at scale and its provision of nuanced understandings of public sentiment through sentiment scores. Its limitations include a dependence on the quality and objectivity of the lexicon used and the difficulty of interpreting nuanced or mixed sentiments.

2.2.3 Surveys

Surveys, in which structured questionnaires are administered to a sample population, have long been a conventional method to gauge public sentiment. Their strengths include the ability to

obtain targeted insights and provision for designing questions to capture specific data. The limitations of surveys are their potential for response biases and their poor scalability and timeliness compared to automated methods.

2.2.4 Relation to Chosen Research Approach

The methodological approach chosen for this study aligns with previous studies that employed machine learning and sentiment analysis techniques. Utilising machine learning for sentiment analysis of Twitter data facilitates the extraction of broad public sentiment trends regarding the COVID-19 pandemic in New Zealand, thereby aligning with the global trend of leveraging machine learning and sentiment analysis in understanding public reactions to the pandemic.

2.3 General Sentiments Relating to the COVID-19 Pandemic around the World

The COVID-19 pandemic started as a natural crisis, but soon the world realised that it was also becoming a social crisis. Every section of society was negatively impacted in different ways. There was a wide range of emotional reactions such as anxiety, fear, depression, future worry, uncertainty, confusion, and so on. In terms of mortality, the COVID-19 fatality rate was about 5% globally (Ahmad et al., 2021). This is less than half the mortality rate of Severe Acute Respiratory Syndrome (SARS), which emerged in 2002 and had a fatality rate of around 11%. COVID-19 is also much less deadly than Middle East Respiratory Syndrome (MERS), way has a fatality rate of 30–40%. However, in terms of public sentiments, the reaction to the COVID-19 pandemic was far more profound than these other health crises.

The reasons for this could be an absence of vaccine treatment for a long time, the speed of spread, a longer incubation time, and the severe adverse impact on the daily lives of adults and children alike (Ahmad et al., 2021). It can now be seen that the impact of the pandemic was almost the same for adults and children. This was contrary to early rumours that children were immune to the severe impact of COVID-19. Initial policy guidelines, both in New Zealand and overseas, created confusion, which led to children and parents taking a very casual approach to dealing with the virus. This raises questions about whether more stringent preventive

measures should have been imposed for children by all governments an area this research will delve further into.

Research published in March 2022 states that the COVID-19 crisis arrived as a tsunami of destructive force and was extremely disruptive. It did not spare anyone, and the emotional distress was felt by all ages, directly or indirectly (Wang et al., 2022). In one case, a son scaled the walls of the hospital to say goodbye to his mother, who was dying of COVID-19 (Jain, 2020). Cases such as this highlight the indirect impact of pandemic on children. If one reflects on these various events, it is easy to realise that the pandemic was devastating for adults and children alike. There is every possibility that some policymakers may have misjudged the novel coronavirus in the absence of scientific research. As such, the guidelines for children for COVID-19 care and compliance may have been wrong in using a different yardstick for adults and children, especially in relation to indirect impacts on various ages.

Turke (2020) asserts that COVID-19's impact appears to be age-related, with younger individuals exhibiting a relative advantage over older ones in terms of disease severity. The study outlines five reasons for this trend, two based on medical grounds and three rooted in the evolutionary theory of life history. Turke (2020), however, urges caution against the overgeneralisation that all age subsets ranging from neonates to young adults possess an inherent advantage over adults in mitigating COVID-19. Turke (2020) notes the lack of unequivocal evidence to substantiate this claim, citing the example of the rotavirus, which significantly impacts infants, causing severe dehydration. This serves as a potent reminder of the age-specific nature of vulnerability to different viruses.

Critiquing the common research trend of focusing on the direct impacts of COVID-19 on children, Turke (2020) also draws attention to the potential oversight of indirect effects. The implementation of policies that suggest children have an advantage in fighting COVID-19 could, as a consequence, inadvertently expose them to unforeseen long-term risks. Turke therefore (2020) underscores the necessity for a comprehensive approach in developing COVID-19 strategies, one that takes into account potential long-term effects on all age categories, and advocates for more nuanced research to fully understand the age-specific impacts of the virus.

In contrast, a study conducted by Roy et al. (2022) in the United States found that COVID-19 affects adults and children similarly. The researchers projected that, by the end of July 2021, both children and adults would need to be vaccinated to protect the most vulnerable and elderly

populations. This strategy was deemed necessary to prevent the health system from being overwhelmed, leading to increased fatalities. Unlike the aforementioned 'youth advantage', this study underscores the necessity of vaccinating all age groups to control the virus spread effectively. Thus, according to these researchers, the 'youth advantage' theory should be considered with scepticism, and the sentiments of both adults and children should be given equal weight in formulating pandemic-related policies (Roy et al., 2022).

In a comprehensive meta-analysis, Romero Starke et al. (2021) questioned the commonly held belief that age significantly influences susceptibility to COVID-19. Their findings argue against policies that link COVID-19 prevention strategies primarily to age, emphasising that such approaches lack substantial evidence. They propose that age-centric preventive measures should give way to risk-based strategies, providing a more accurate and effective framework for managing the pandemic (Romero Starke et al., 2021).

2.4 Adult Sentiments on Twitter Relating to the Pandemic in New Zealand and around the Globe

Social media has served as a barometer for public sentiment during the COVID-19 pandemic. Twitter experienced a surge of pandemic-related discourse in the early months of 2020, following WHO's formal announcement that COVID-19 had reached the status of global pandemic. These tweets, totalling nearly 20 million between January and April 2020, encapsulated a broad spectrum of emotions, from fear and anger to joy and sadness. The fluctuations in public sentiment, experienced similarly by adults and children, highlighted the pandemic's pervasive effects on mental well-being, as communities grappled with isolation and loss (Lwin et al., 2020).

Research by Lwin et al. (2020) underscores the need for effective public health communication. Without it, public mistrust and scepticism may proliferate, complicating efforts to combat the virus (Lwin et al., 2020). Given that emotions such as fear, anger, joy and sadness were universally experienced, irrespective of age, these findings demonstrate the importance of policymakers considering the psychological impacts of health crises when designing interventions.

Lwin et al. (2020) analysed over 105 million tweets globally, covering France, Italy, China, the United Kingdom, the Arab region and Spain. These countries together represent a population of approximately 2.4 billion, and the goal was to better understand the global sentiment and inform

effective policy interventions. The study found similar emotional responses across these diverse regions, with expressions of chaos, disaster, anxiety, denial, sadness and fear proliferating on the platform (Lwin et al., 2020). Interestingly, age was not a significant factor in the expression of these sentiments, hinting at universal experiences under the shared circumstances of the pandemic.

In a subsequent study, Lwin et al. (2022) scrutinised Twitter posts from the United States, the United Kingdom, India, Singapore and South Korea to compare public sentiment during the pandemic. This research revealed a surge in negative sentiment across all examined countries. However, there was also a concurrent increase in positive emotions, indicating resilience, support and hope among the public (Lwin et al., 2022).

A national-level study by Jafarzadeh et al. (2021) focused on Twitter discussions about COVID-19 in New Zealand. This research highlighted how social media, particularly Twitter, allows individuals to communicate their experiences and expectations during rapidly evolving crises. Analysing these online discourses offers a vital tool for policymakers to understand public sentiment and develop effective strategies. Jafarzadeh et al. (2021) analysed a large number of tweets, primarily focusing on New Zealanders' experiences and responses to the pandemic, without distinguishing between adults and children.

Finally, a study by Ali and Liu (2021) noted the reliance of adults on various communication channels during the pandemic. The authors suggest that regular, interactive communication from government bodies, such as the Ministry of Business, Innovation and Employment (MBIE) in New Zealand, is essential to mitigate public uncertainty. However, it remains unclear whether these communications effectively differentiated between adults and children (Ali & Liu, 2021).

These studies collectively suggest that people's emotional responses to the pandemic on Twitter were universally shared across different age groups and geographical locations. While these findings offer valuable insights for policymakers, they also indicate a research gap in understanding age-specific responses to the pandemic. This warrants further investigation.

2.5 The Impact of the Pandemic on Young People in New Zealand and the World

As the magnitude of information shared on mainstream and social media continues to expand, a well-informed selection of its features and variables increases the precision of sentiment

analysis. An increasing number of policymakers are now exploiting various social media platforms and surveillance methods. In particular, Twitter has proven to be a valuable resource to gauge public sentiment, both in general and specific contexts.

Governments and other entities are increasingly utilising Twitter data to devise appropriate strategies at both the macro and micro level. In line with this trend, research is being undertaken to understand the pandemic-related sentiments of children active on Twitter in New Zealand and globally (Srikanth et al., 2022). This research aims to discern whether pandemic-related sentiments among adults and children were homogeneous or diverse, which could inform the creation of customised, effective, COVID-related policies specifically tailored for children in New Zealand. The youth active on Twitter and other social media platforms constitute a significant demographic that should not be disregarded as 'low risk' due to their assumed high immunity. Sentiment analysis of their online communication is crucial in addressing their unique issues and involving them in combatting the virus if necessary (Srikanth et al., 2022).

The impact of COVID-19 on youth is explicitly laid out in a report published by the Organisation for Economic Co-operation and Development (OECD, 2020). The report highlights that the virus has severely affected some young people, especially the poor, homeless and detained. Policy decisions such as school closures and social distancing have disrupted children's routines and indirectly increased instances of malnutrition and domestic violence. Therefore, careful monitoring of feedback from children across various platforms, including real-time platforms like Twitter, is essential. These insights can guide immediate policy changes to address issues such as increased sexual exploitation and cyberbullying due to unsupervised internet usage during school closures.

These observations suggest that policymakers can formulate customised policies based on the challenges faced by children in various parts of the world. There's an evident gap between the information available on Twitter and other social media platforms and the reality on the ground, making some policies impractical for addressing children's issues. Therefore, understanding the needs of different societal segments, such as adults and young people, is crucial (OECD, 2020).

In another significant study, it was revealed that the stress related to COVID-19 experienced by pregnant mothers can potentially affect their unborn baby's brain development, which can have long-term implications on their mental and physical skills (Wenner Moyer, 2022; see also

Han et al., 2021). This realisation contradicts the prevailing belief that infants naturally cope well with COVID-19. This and similar assumptions may be responsible for policy decisions focusing predominantly on the adult population (Wenner Moyer, 2022).

Recent research studies suggest that COVID-19 case numbers among children have been significantly underestimated due to unreported cases. By February 2022, children infected with COVID-19 accounted for 75% of all children in the 0–17 age group (Clarke et al., 2022). Contrary to the belief that young people have a lower risk of infection, surveillance research confirmed that infection rates in children are comparable to adults. Surprisingly, studies have found higher pathogen levels in children under 18 compared to adults (Deville et al., 2022). Reported cases in the United States alone stood at nearly 16 million for the age group below 18 years (Deville et al., 2022).

Overall, the COVID-19 pandemic's influence on young people, both within New Zealand and on a global scale, underscores the importance of dedicated and accurate research. Conventional wisdom or broad assumptions about youth resilience or immunity must give way to rigorous investigation and nuanced understanding. Sentiment analysis of young people's online communication on platforms such as Twitter can yield valuable insights that could aid in crafting effective, tailored strategies and policies to mitigate the pandemic's impact. The data paints a clear picture: children and young people are not merely bystanders in this global health crisis but active participants, warranting careful consideration and targeted response from policymakers and society at large.

2.6 Comparative Analysis of Public Sentiments across Countries during the COVID-19 Pandemic

The COVID-19 pandemic precipitated a wide range of public sentiments around the world. The diverse socio-political landscapes, healthcare systems, and governmental policy responses have significantly influenced public reactions. This comparative analysis aims to elucidate the similarities and differences in public sentiments across various nations during the pandemic, employing studies that leveraged social media platforms as a window to gauge public emotions and perceptions.

Following the outbreak of the pandemic, a surge in the use of social media platforms for expressing emotions was seen globally, which displayed a significant decline in sentiment, (Dizikes, 2022; Lwin et al., 2022). The utilisation of social media as a platform for expressing pandemic-related information was notably observed in countries like India, Singapore, South Korea, the United Kingdom, and the United States (Lwin et al., 2022; Yu et al., 2022).

Yu et al. (2022) highlighted the differing emotional reactions across India, Singapore, South Korea, the United Kingdom, and the United States, with each nation exhibiting unique sentiment trajectories correlated with the timing and nature of governmental actions in response to the pandemic (Lwin et al., 2022). Additionally, the rate of sentiment recovery varied across different regions, as demonstrated by Wang et al. (2021), which was attributed to the efficacy of governmental interventions and economic stimulus packages across different nations. The contrasting public sentiments underscore the complex interplay between public sentiment, governmental responses, and societal dynamics during the COVID-19 pandemic across different nations.

This comparative analysis illuminates the multifaceted nature of public sentiments during the COVID-19 pandemic around the globe. A nuanced understanding of the complex interplay between public sentiment, governmental responses, and societal dynamics is pivotal for crafting effective, culturally sensitive, and informed policy interventions on both the national and global scales. The findings underscore the imperative for a globally informed approach in policy formulation and crisis management, accentuating international cooperation and shared learning in navigating the challenges posed by a global pandemic.

2.7 Conclusion

The findings of this literature review underscore the critical role of social media platforms, and specifically Twitter, as a valuable source of public sentiment during the COVID-19 pandemic. The studies reviewed above reveal a striking similarity in emotional responses across different age groups and national boundaries, pointing to the universality of human reactions during a crisis.

The meta-analysis by Romero Starke et al. (2021) challenged the assumption that age is a determining factor in COVID-19 infection risk. Concurrently, the research conducted by Lwin et al. (2020), Lwin et al. (2022) and Jafarzadeh et al. (2021) underscores the importance of properly managed public communication in times of crisis. These studies demonstrated that the

sentiments expressed by individuals on Twitter offer a real-time barometer of public mood and sentiment, which is increasingly being utilised by policymakers.

Moreover, the literature review brought attention to a significant gap in the understanding and inclusion of the sentiments of young people regarding the pandemic. As Srikanth et al. (2022) emphasised, young people form a vital demographic whose sentiments are equally important to consider when formulating health policies and strategies.

Various studies, including those by Wenner Moyer (2022), Han et al. (2021) and Deville et al. (2022), highlighted the need for age-specific research and interventions. They demonstrate that the pandemic's impact on children and young people has been underestimated and under-reported, and call for more specific research and distinct, tailored policy approaches.

In summary, the literature suggests that a one-size-fits-all approach is unlikely to be effective in managing a crisis as complex and far-reaching as the COVID-19 pandemic. The need for nuanced, evidence-based, and demographic-specific strategies is clear. Further research, particularly focusing on the experiences and needs of young people, is necessary to develop more comprehensive and effective public health policies and interventions.

Chapter 3. Methodology

3.1 Introduction

This chapter outlines the methodology used in this research, which encompasses research design, data collection and analysis, and how the data is presented. In addition, the potential limitations of the analysis are discussed, and the ethical considerations are addressed. As explained below, this research adopted a mixed-methods approach, incorporating both qualitative and quantitative analyses, to investigate trends and patterns in public sentiment regarding COVID-19 vaccination among New Zealanders. To achieve this, the research employs sentiment analysis techniques using machine learning algorithms to analyse tweets by people living in New Zealand.

3.2 Research Design

A research methodology serves as a comprehensive guide to data collection and analysis techniques, explaining the researcher's choice of methods in conducting the study. It communicates the steps taken to achieve the research aim and details the 'how' aspect of reaching the research goal. By providing this information, readers of the research can better assess the validity and reliability of the results (McCombes, 2023).

This research employed a mixed-methods approach, incorporating both quantitative and qualitative data collection. Quantitative analysis is a research method that utilises numerical data to quantify, interpret and analyse relationships and patterns in a systematic manner. This type of analysis is essential for the generalisation of findings, hypothesis testing, and predicting outcomes in various fields, such as the social sciences, economics, psychology and natural sciences (Creswell, 2014).

Quantitative data consists of numerical information that can be measured and counted. In the case of the present study, such information includes the number of retweets, likes and replies associated with a tweet. This data can help assess a tweet's popularity and identify trends in public opinion.

Conversely, qualitative data focuses on non-numerical information, encompassing the content and context of the tweets. This type of data captures the emotions, opinions and ideas expressed within the text. Sentiment analysis leverages qualitative data to classify tweets as positive,

negative, or neutral based on the words and phrases used. By integrating quantitative and qualitative data, this research aims to provide a comprehensive understanding of public sentiment regard COVID-19 vaccination, enabling the examination of both the popularity and emotional nuances of tweets.

3.3 Data Collection from Twitter

After the scandal of the massive data privacy breach involving Cambridge Analytica that erupted in 2018 (Criddle, 2020), it has become difficult to retrieve data for research from many social media platforms. Most platforms now have restricted access to their Application Programming Interface (API). Fortunately for the purposes of this research, Twitter continues to provide access to its API. Therefore, data collection was carried out via Twitter, which is being increasingly utilised in academic research.

Since the focus of this research is on impacts, this formed the framework and context for accessing Twitter. The data was collected directly from Twitter by observing the sentiments of an online community affected by COVID-19. This approach combines elements of 'netnography' and digital research techniques to study online communities and their cultural practices. Originating from the work of Robert Kozinets, netnography involves the systematic collection and analysis of digital data, focusing on the interactions and conversations of individuals within virtual spaces (Kozinets, 2015). Also known as 'digital ethnography', this type of research is a form of specialised qualitative research (Gleeson, 2022). This approach enables researchers to gain a deeper understanding of online social dynamics, behaviours and sentiments, making it particularly valuable for examining contemporary issues and phenomena in the digital age.

Digital ethnography repurposes traditional ethnographic practices for studying online environments and digital interactions. It involves the observation, documentation and analysis of behaviours, attitudes and conversations of individuals within digital spaces (Hine, 2015). This approach allows researchers to immerse themselves in the online world, similar to how anthropologists embed themselves within the physical communities they are studying (Murthy, 2008).

In the context of our increasingly digital lives, digital ethnography is an essential instrument to probe human interaction and experiences within online communities. As emphasised by Pink et

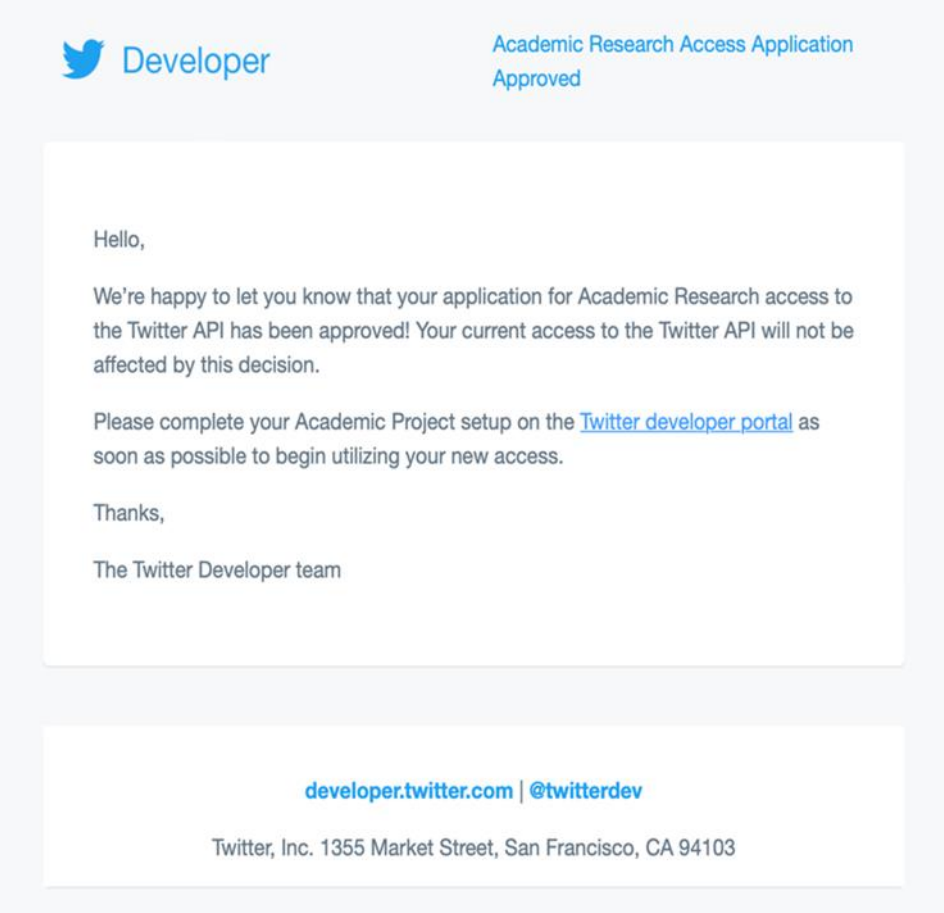
al. (2016), this research approach acknowledges the intertwined relationship between online and offline worlds, viewing them as interconnected realms rather than separate entities. It permits the researcher to capture the richness and complexity of social relations, culture and experiences that emerge in these digitally mediated environments.

Furthermore, digital ethnography recognises the variety of digital platforms and channels, along with the various ways individuals use them (Postill & Pink, 2012). As such, the analysis transcends text-based communication, considering a range of multimedia forms including images, videos, emojis, hashtags and hyperlinks. This multi-faceted approach offers a more comprehensive understanding of the communicative practices and cultural norms within the digital environment.

The dataset for this study was gathered through Twitter's official academic API. The reason for selecting Twitter was the wide variety of digital ethnography tools that are compatible with it. The tweets posted by New Zealand users on Twitter between 1 February 2021 and 27 September 2022, specifically those relating to COVID-19 vaccination, were mined for data. To extract the tweets, a Twitter developer account was required. The information regarding the Twitter API and keys was saved in variables, and an authentication object was constructed; an access token was then obtained. Dr Abdulaziz Alaboudi, who holds a PhD in Software Engineering, provided valuable guidance in making an appropriate decision regarding the selection of research tools (Ahmed, 2021). Recently, the ethics of social media research has gained academic attention. Hence, in deciding to use Twitter, ethical, legal and methodological factors were all considered (Ahmed et al., 2017).

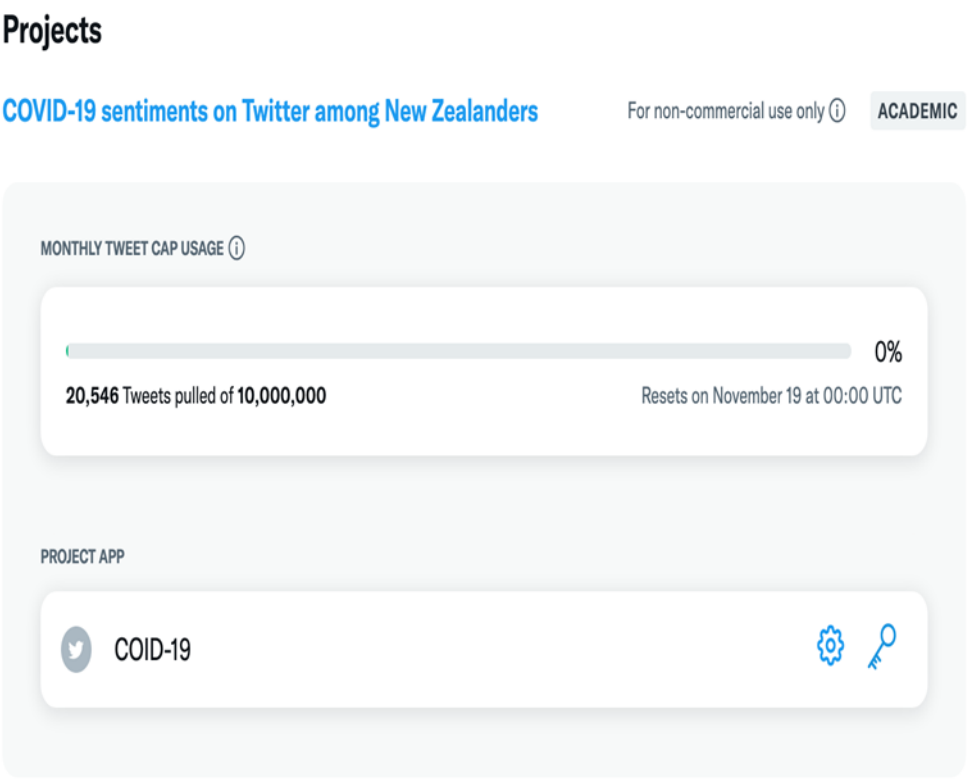
Figure 2

Twitter academic approval email



The first step in the data collection process involved using an existing Twitter account to apply for developer access, which took two days to receive approval (Figure 2). A project was then created inside the API account and named 'Sentiments on Twitter among New Zealanders'. This API allows for more than 330 million tweets per month (Figure 3).

Figure 3
The Twitter project inside the Twitter Academic Portal



In this study, a Python script incorporating geotagging was developed to filter tweets originating from a specific location, in this case New Zealand. A Python script with geotagging is a piece of code written in the Python programming language that is designed to work with geospatial data by utilising geographic coordinates such as latitude and longitude. Geotagging refers to the practice of embedding geographic information within various digital content forms, including images, videos, or social media posts.

Within the context of a Python script, geotagging often entails parsing, analysing, or visualising data containing geographic information to derive insights or meaningful conclusions based on the geospatial context. This process was completed, and a screenshot of the results appears as Figures 7-14 in Chapter 4 on results and analysis.

The Python script focused on identifying vaccine-related sentiments from 1 February 2021 to 27 September 2022, targeting both adult and child users. Data collection for this study involved using specified queries and hashtags on the social media platform Twitter.

3.4 Tweet Inclusion and Exclusion Criteria

Tweets were collected if they included one or more of the following hashtags: #COVID-19, #CovidVaccine, #adultvaccine, #childrenvaccine, #Pfizer and #AstraZeneca.The inclusion of these hashtags in the study was effective due to their direct relevance to the research topic, which focused on public sentiment regarding COVID-19 vaccinations. These specific hashtags are widely used by individuals when discussing COVID-19 vaccines on social media platforms like Twitter, ensuring that a larger and more diverse dataset can be collected for analysis. Furthermore, using these targeted hashtags allows for a more focused and precise data collection process, filtering out unrelated content and conversations. This precision helps in obtaining a higher quality dataset, which in turn leads to more accurate findings and conclusions. The results were read and loaded into a dataset for further analysis. An excerpt of a Python script used to collect data related to the COVID-19 vaccine for kids is shown in Figure 4.

To identify relevant tweets, a systematic approach was taken utilising specific keywords and phrases. The main criteria for filtering tweets were:

1. Location and language: The tweets were filtered to come from New Zealand (‘place_country:NZ’) and were written in English (‘lang:en’).
2. Vaccine keywords: The search prioritised tweets discussing various vaccine-related terms such as ‘vaccine’, ‘boosters’, ‘AstraZeneca’, ‘Moderna’, ‘Sinovac’, ‘vaccination’, ‘Pfizer’, and ‘unvaccinated’.
3. Age-related keywords: To specifically target tweets related to children and teenagers, a comprehensive set of age-related terms were included. These comprised general terms like ‘teenage’, ‘child’, ‘children’, ‘teenagers’, ‘young’, ‘youth’, ‘son’, and ‘daughter’.
4. Explicit age criteria: The query also included specific age mentions to capture a wider range of tweets that might indicate the age of children. These ranged from ‘1 year old’ up to ‘15 years old’.
5. Iterative time-frame search: The script fetched tweets iteratively based on specific date ranges (the ‘dateVac’ variable in Figure 4). For each date in this range, tweets were fetched between the start time of that date at 01:00:00 and its end time at 28:00:00, ensuring a broad capture window.
6. Consolidation of data: After fetching tweets based on the criteria, they were structured into a Data Frame to organise and manage the data more efficiently. The ‘text’ of the tweet was extracted and stored along with the respective date.

7. Categorisation: Lastly, these filtered tweets were assigned a category label of ‘kids’ for further analysis and differentiation.

This methodical approach ensured a comprehensive and targeted data collection, prioritising both the topic of interest (vaccination) and the demographic in focus (children).

Figure 4

A Python script used to collect data related to the COVID-19 vaccine for kids

```python

from time import sleep


tweets_clean_kids = pd.DataFrame()

query = '(place_country:NZ lang:en -is:retweet) (vaccine OR boosters OR AstraZeneca OR Moderna'\
 'OR Sinovac OR vaccination OR Pfizer OR unvaccinated) (teenage OR child OR children OR teenagers'\
    'OR young OR youth OR son OR daughter OR 1 year old OR 2 year old OR 3 year old OR 4 year old OR 5 year'\
    'old OR 6 year old OR 7 year old OR 8 year old OR 9 year old OR 10 year old OR 11 year old OR 12 year'\
    'old OR 13 year old OR 14 year old OR 15 year old OR 1 years old OR 2 years old OR 3 years old OR 4 years'\
    'old OR 5 years old OR 6 years old OR 7 years old OR 8 years old OR 9 years old OR 10 years old OR 11 years old'\
    'OR 12 years old OR 13 years old OR 14 years old OR 15 years old)'

for date in dateVac:
    rawTweets = api.search_all_tweets(query, start_time=f'{date}-01T00:00:00Z', end_time=f'{date}-28T00:00:00Z', max_results=500)
    if rawTweets[0] is None:
        continue
    tweets_text = [tweet['text'] for tweet in rawTweets[0]]
    newPd = pd.DataFrame(tweets_text, columns=['text'])
    newPd['date'] = date
    tweets_clean_kids = pd.concat([tweets_clean_kids, newPd])
    print(f'Finished {date}')
    sleep(10)

tweets_clean_kids['age'] = 'kids'
```

3.5 Data Processing and Cleaning

This research was primarily focused on gauging the sentiment of adult and young New Zealanders regarding the country’s COVID-19 vaccination programme. During the quantitative data collection process, data measurements were straightforward. Converting fuzzy sentiment variables into measurable indicators was achieved by following the guidelines set by McCombes (2023).

For the sentiment analysis, the Twitter-roBERTa-base was employed This machine learning model has been specifically fine-tuned for analysing sentiments in tweets. The underlying architecture is RoBERTa, a derivative of the BERT (Bidirectional Encoder Representations from Transformers) model. RoBERTa was developed by Facebook AI and has led to notable improvements in a multitude of natural language processing (NLP) tasks, especially sentiment analysis.

RoBERTa was comprehensively trained on vast datasets, including the BooksCorpus (800 million words) and the English version of Wikipedia (2,500 million words). The training process discarded the next-sentence prediction objective used in BERT. Instead, RoBERTa optimised its process by using larger mini-batches, training with more data overall, and incorporating dynamic masking. With regard to validation, RoBERTa was subjected to several benchmarks in NLP, ensuring its robustness and reliability in a variety of tasks.

Although RoBERTa offers an impressive generic foundation for sentiment analysis, tweets, given their nature, present unique challenges. They're concise, replete with colloquialisms, abbreviations, and cultural nuances. Recognising these intricacies, this research didn't attempt to fine-tune RoBERTa specifically for COVID-19 vaccination tweets. The reason for this is twofold:

1. **Broad and comprehensive training:** RoBERTa''s extensive training data already encompasses diverse linguistic patterns and sentiments, offering a strong foundation for understanding the sentiment of tweets, even without domain-specific fine-tuning.
2. **Resource consideration**: Given RoBERTa's robust pre-trained capabilities, fine-tuning, which requires significant computational resources and a well-structured, domain-specific database, wasn't deemed necessary.

Before delving into analysis, the tweets underwent rigorous preprocessing and cleaning to ensure they were optimally primed for sentiment analysis. The procedure involved the following steps:

Text preprocessing: The custom 'preprocess(text)' function was employed to meticulously scour and sanitise the tweet content.

Usernames starting with '@' were filtered out to prevent any potential bias stemming from user mentions.

Next, URLs that began with prefixes such as http or https were expunged to ensure the sentiment wasn't skewed by external links.

Finally, the keywords ‘vaccine ‘, ‘Vaccine‘, ‘vaccines‘, ‘Vaccines‘, ‘vaccination‘, and ‘Vaccination‘, were neutralised. This action ensured the analysis honed in on the context surrounding vaccination, rather than repetitively flagging these terms.

1. **Tokenisation**: Post-cleaning, the tweets were tokenised using the ‘AutoTokenizer.from_pretrained(MODEL)’ function from the transformers library. This crucial step transformed the cleaned text into a model-friendly format, breaking the text into consumable tokens or subwords which were then mapped to their respective IDs for the RoBERTa model to process.
2. **Noise data minimisation**: Emphasis was placed on excluding noise data – irrelevant words, typographical errors, special characters, and superfluous URLs. Such data, if unchecked, can compromise the analysis, introducing inaccuracies and inefficiencies. By diligently removing these potential pitfalls, the research was better poised to extract genuine sentimental insights from the tweets.

By integrating this comprehensive cleaning and preprocessing procedure, the research ensured that the sentiment analysis was both precise and dependable, focusing intently on the tweets" core sentiment while minimising external influences.

The sentiment analysis model was presented with tweets as plain text, deliberately excluding IDs to guarantee result anonymity. Upon analysis, each tweet was categorised as either positive, negative, or neutral. The ultimate classifications were stored in an Excel document, forming the basis for subsequent result discussions. A comprehensive breakdown of the dataset analysis is illustrated in Figure 5.

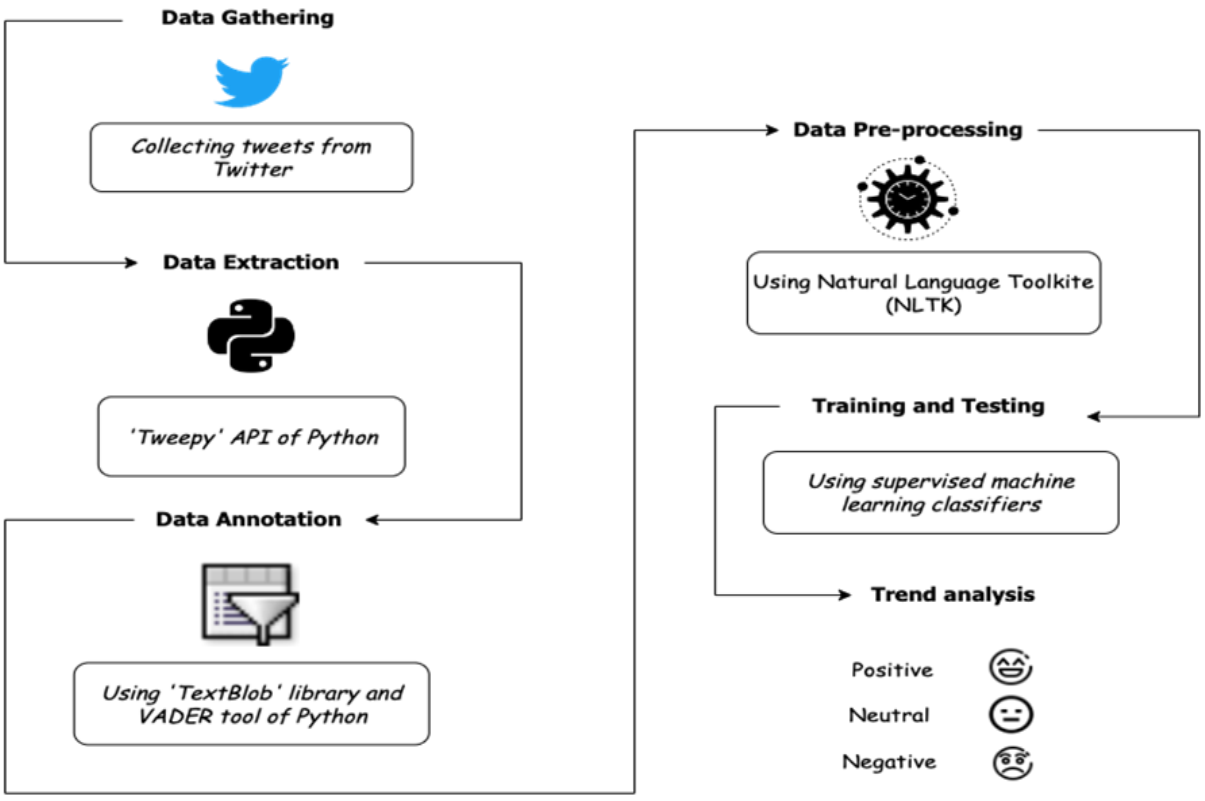
Figure 5

The Python script used for the sentiment analysis using Twitter-roBERTa-base

```python
from transformers import AutoModelForSequenceClassification
from transformers import TFAutoModelForSequenceClassification
from transformers import AutoTokenizer, AutoConfig
import numpy as np
from scipy.special import softmax
# Preprocess text (username and link placeholders)


def preprocess(text):
    new_text = []
    for t in text.split(" "):
        t = '' if t.startswith('@') and len(t) > 1 else t
        t = '' if t.startswith('http') else t
        t = '' if t.startswith('https') else t
        t= '' if t.startswith('vaccine') else t
        t= '' if t.startswith('Vaccine') else t
        t= '' if t.startswith('vaccines') else t
        t= '' if t.startswith('Vaccines') else t
        t= '' if t.startswith('vaccination') else t
        t= '' if t.startswith('Vaccination') else t
        new_text.append(t)
    return " ".join(new_text)


MODEL = f"cardiffnlp/twitter-roberta-base-sentiment-latest"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
config = AutoConfig.from_pretrained(MODEL)
# PT
model = AutoModelForSequenceClassification.from_pretrained(MODEL)
#model.save_pretrained(MODEL)
tweets_analysis = []
for index, row in allTweets.iterrows():
    text = row['text']
    text = preprocess(text)
    encoded_input = tokenizer(text, return_tensors='pt')
    output = model(**encoded_input)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    # Get the sentiment with the higher score
    ranking = np.argsort(scores)
    ranking = ranking[::-1]
    top_sentiment = config.id2label[ranking[0]]
    tweets_analysis.append(
        {
            'text': text,
            'sentiment': top_sentiment,
            'date': row['date'],
            'age': row['age'],
            'date': row['date']
        }
    )
    print(f'Finished {index} {top_sentiment}')

```

3.6 Sentiment Analysis

Figure 6
Framework for COVID-19 vaccination sentiment analysis using Twitter content



Descriptive statistics were then applied in terms of the distribution of the source material and the mean or central tendency interpreted to identify trends and patterns among the sentiments.

In summary, the following six steps were followed during the data collection and analysis (see Figure 6):

1. **Data gathering**: collecting relevant tweets using queries, hashtags and geotagging.
2. **Data extraction**: extracting useful information, such as text and metadata, from the collected tweets.
3. **Data annotation**: labelling the extracted data with sentiment values, typically positive, negative, or neutral.
4. **Data pre-processing**: cleaning and formatting the data to remove noise and prepare it for analysis.

5. **Training and testing**: using the machine learning model Twitter-roBERTa-base to learn from the labelled data and evaluating its performance.
6. **Trend analysis**: a comprehensive research technique used to identify, analyse and interpret patterns and shifts in data over time. Trend analysis entails a detailed examination of the data output in this case the sentiments expressed in Twitter data regarding COVID-19 vaccinations among New Zealanders to identify discernible patterns, significant shifts, and notable insights. The aim is to understand the trajectory of these sentiments whether they're positive, negative, or neutral across various timeframes and demographic groups.

This analytical process was not solely focused on the quantitative aspects but also included qualitative dimensions, which involves more than just categorising data into time periods or assigning sentiment scores based on a machine learning algorithm's output. It involves generating visual representations to better comprehend changes in public sentiment over time and delving into specific demographic groups to understand sentiment trends within these cohorts.

Moreover, trend analysis employs thematic analysis to identify recurring themes, topics, or concerns in the tweets. This helps gain a deeper understanding of the factors influencing public sentiment, providing a richer, more nuanced context to the sentiment scores. Therefore, trend analysis serves as a holistic approach, integrating both qualitative and quantitative elements to provide a comprehensive understanding of public sentiment trends regarding COVID-19 vaccinations. The insights gathered through this method can inform public health communication strategies, tackle vaccine hesitancy, and aid in improving vaccination rates.

Table 1 shows examples of the kinds of Tweets included in the analysis for this research.

Table 1
Examples of Tweets collected

Tweets	**Date**	**Group**
Luckily we are in front of the queue. #nzpol #COVID19nz	2021-02	Adult
"If the doesn't allow us to resume normal life then WHATS THE POINT OF THE VACCINE"	2021-02	Adult
So why are you promoting anti covid themes? There is no global conspiracy only a global pandemic. Humans have worked tirelessly to produce safe, effective C19 to prevent loss of life & damage from the virus. Have you ever celebrated this amazing outcome?	2021-02	Adult

"In one example, the experts concluded that Pfizer had deliberately delayed release of a study that showed the drug (gabapentin) had little effect against the pain that is a complication of long-term diabetes" Experts conclude that Pfizer manipulated studies.....	2021-05	Adult
No side effects yet from the second 😊☺	2021-05	Adult
His point is valid though, and how I also interpreted your first reply. If your child has had bad reactions to previous that doesn't mean that would be the case with an mRNA**,** Have you discussed this with your GP? Regardless, your tweet does have strong anti-vax vibes	2021-12	Kids
You are aware that children aren't just simply small adults right? Comparing pregnant people and children is a false equivalence. Just because a treatment/vaccine/intervention/drug gets approved elsewhere, means we should just dispense with the whole process.	2021-12	Kids
WHO says no evidence healthy children, adolescents need COVID-19 boosters - Reuters https://t.co/dytvD1n9sr	2022-01	Kids
Got my 9-year-old son vaxxed today. Amazing operation as usual at the Tamaki Centre.	2022-01	Kids
Coronavirus: Pfizer/BioNTech COVID less effective in children aged 5-11 - study	2022-03	Kids

3.7 Limitations of the Analysis

The are a number of limitations of sentiment analysis, with one of the most prominent relating to its ability to identify phrases that accurately encapsulate the sentiment of a given text. The intricacies of NLP can pose significant challenges in analysing sentiment or emotions. This is largely due to the fact that computers require extensive training to be able to interpret and comprehend emotions in a manner akin to human cognition.

As technology and data science evolve, sentiment analysis tools are expected to become increasingly proficient at addressing these challenges. Among these challenges, three key limitations are the potential for inaccurate interpretation, temporal snapshot constraints and inherent bias (Islam, 2020).

Inaccurate interpretation arises from the use of idioms, emojis and sarcasm. By default, machine learning algorithms, interpret things in their literal sense, which makes understanding figures of speech based on idioms difficult. Consequently, an idiom used in a comment or review may be misinterpreted or entirely overlooked by the algorithm. This issue is further exacerbated when multiple languages are involved, such as English and te reo Māori.

Emojis present another unique challenge, particularly for text-based social media platforms like Twitter where emoji usage is prevalent. NLP tasks are typically designed to be language-specific, which complicates emoji interpretation.

Finally, irony and sarcasm, which are prevalent in casual conversation and social media memes, pose significant challenges for sentiment analysis. Negative sentiments conveyed through backhanded compliments can be difficult for sentiment analysis techniques to accurately interpret. This tends to result in a disproportionate representation of positive feedback compared to negative feedback.

In conclusion, emojis, sarcasm and idioms can potentially hinder the accuracy of sentiment analysis, leading to skewed or inaccurate interpretations of the data.

3.8 Methodological Considerations: Absence of Statistical Tests

The main objective of this research is to explore and analyse trends in public sentiment towards New Zealand's vaccination programme, with a dual focus on adults and children. Consequently,

the methodological approach needed to be rooted in exploratory data analysis (EDA) as opposed to hypothesis testing, leading to an absence of statistical tests in the analysis.

Exploratory data analysis is a research approach that allows us to understand the data by summarising its main characteristics, often with visual methods (Tukey, 1977). It is often used when a researcher doesn't have a clear idea of what patterns or relationships the data might reveal. In contrast, hypothesis testing starts with a specific assumption or hypothesis about the data that is then tested for statistical significance.

In the context of this study, the use of EDA is beneficial for several reasons. First, it allows us to openly investigate the dataset without making any initial assumptions, an approach suitable when dealing with a novel and multifaceted issue such as public sentiment during a pandemic (Hastie et al., 2009). Second, it provides the flexibility to analyse and compare sentiments across diverse demographic groups, including both adults and children, without being confined to a pre-determined hypothesis. Finally, EDA is particularly effective when coupled with machine learning techniques, enabling us to detect complex patterns and relationships within the large-scale text data derived from digital platforms (Grimmer & Stewart, 2013; Nelson, 2020).

Importantly, the lack of statistical tests does not mean that this analysis is devoid of rigour. The robustness of the findings is ensured through careful data collection, meticulous data processing, and the use of well-established machine learning techniques. The findings are further corroborated by situating them within the broader literature and previous research findings.

Thus, the methodological approach adopted in this research aimed to provide an in-depth and nuanced understanding of public sentiment and to reveal valuable insights that can inform health communication strategies and contribute to the broader acceptance of vaccination within the community (Betsch et al., 2018; Brewer et al., 2017).

3.9 Ethical Considerations

Conducting research using data from online platforms such as Twitter necessitates stringent adherence to ethical guidelines, given the sensitive nature of personal information and the platform's policies. For this study, while direct human subject interaction was absent—eliminating the need for traditional ethical approvals—the focus was firmly on respecting data privacy, user protection, and adherence to legal standards, specifically the General Data Protection Regulation (Regulation 2016/679).

The data handling processes were designed to comply with GDPR mandates, ensuring the anonymization of sensitive information and establishing that data utilization was strictly for research purposes (Floridi, 2016). This anonymization process protected user identities and complied with Twitter's terms of service and community guidelines, which forbid the unauthorized sharing of confidential information (Twitter, 2021).

Further, the research methodology was crafted in line with the ethical considerations proposed by Zimmer and Proferes (2021), recognizing the importance of user consent, data minimization, and the ethical responsibility researchers carry when dealing with digital footprints. These measures were crucial in maintaining the confidentiality and integrity of the extracted data, thereby upholding the users' autonomy and privacy online.

In adherence to these best practices, the research avoided potential ethical pitfalls related to digital data collection and analysis, such as the inadvertent exposure of user identities or biases in data handling. This was particularly relevant given the public nature of the data and the ease with which digital information can be misinterpreted or misused (Flicker et al., 2019).

The study's ethical stance underscored a commitment to contributing to the academic community without compromising the digital rights and privacy of the individuals whose data was instrumental in this scholarly pursuit.

3.10 Summary

This research investigated public sentiment towards COVID-19 vaccination in New Zealand by leveraging a mixed-methods approach with Twitter data. Using Twitter's API, relevant tweets were gathered based on specific hashtags related to COVID-19 vaccines. These tweets were then cleaned to remove irrelevant information and analysed using a machine learning model, Twitter-roBERTa-base, which classified the sentiment of each tweet as positive, negative, or neutral. Despite some limitations like potential biases in the Twitter user base and misinterpretations due to sarcasm or cultural nuances, the study has been designed to provide meaningful insights into the public opinion on COVID-19 vaccinations within the country's digital sphere.

Chapter 4. Findings

4.1 Introduction

The aim of this study was to investigate and analyse the sentiments of adults and young people on Twitter in New Zealand during the COVID-19 vaccination campaign. Detailed data were pulled in an ethical way from Twitter, as explained in the previous chapter. The raw data were summarised to compare the sentiments of the adults and the young people. Next, the focus was on the respective sentiment analyses of the adult and youth tweets. The tweets were categorised into one of three categories: positive, negative or neutral.

4.2 Data Summary

Table 2
Classification of sentiments by adults and kids

GROUP	TWEETS	PROPORTION	NEGATIVE	NEUTRAL	POSITIVE
ADULTS	5909	96%	2430	2556	923
KIDS	262	4%	113	107	42
TOTAL	6171	100%	2543	2663	965

Understanding how people feel about vaccination can inform the decisions of public health professionals and policymakers in health crises. In order to promote and increase vaccination coverage, public health professionals may also be prepared to utilise Twitter and other social media platforms to increase positive messaging, reduce negative and oppositional propaganda, and proactively delete anti-vaccination profiles like bots.

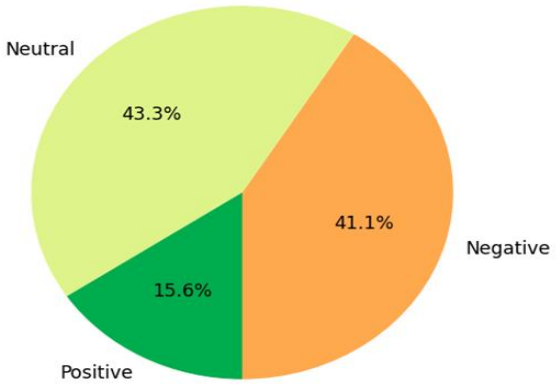
A total of 6171 tweets were collected between February 1, 2021, and September 27, 2022. Table 2 displays the findings of a sentiment analysis conducted on tweets about COVID-19 vaccines for adults and kids in New Zealand. According to the table, 923 tweets for the adult group were rated positively, 2556 were rated neutrally, and 2430 were rated negatively. There were 42 tweets classified as positive, 107 as neutral, and 113 as negative for the children's group. For the adult

group, the percentages of positive, neutral, and negative tweets were 96%, 4% and 0%, respectively, while for the child group, they were 16%, 41% and 43%, respectively. There were 5909 tweets in the adult group and 262 in the children's group.

4.3 Vaccine Sentiment Analysis for Adults

Figure7

Distribution of adult sentiment relating to the COVID-19 vaccine



The graph above shows that 43.3% of adult tweets expressed neutral sentiments towards the impact of the COVID-19 vaccine. This neutrality is significant as it suggests that a large section of the adult population abstained from expressing either positive or negative viewpoints on Twitter.

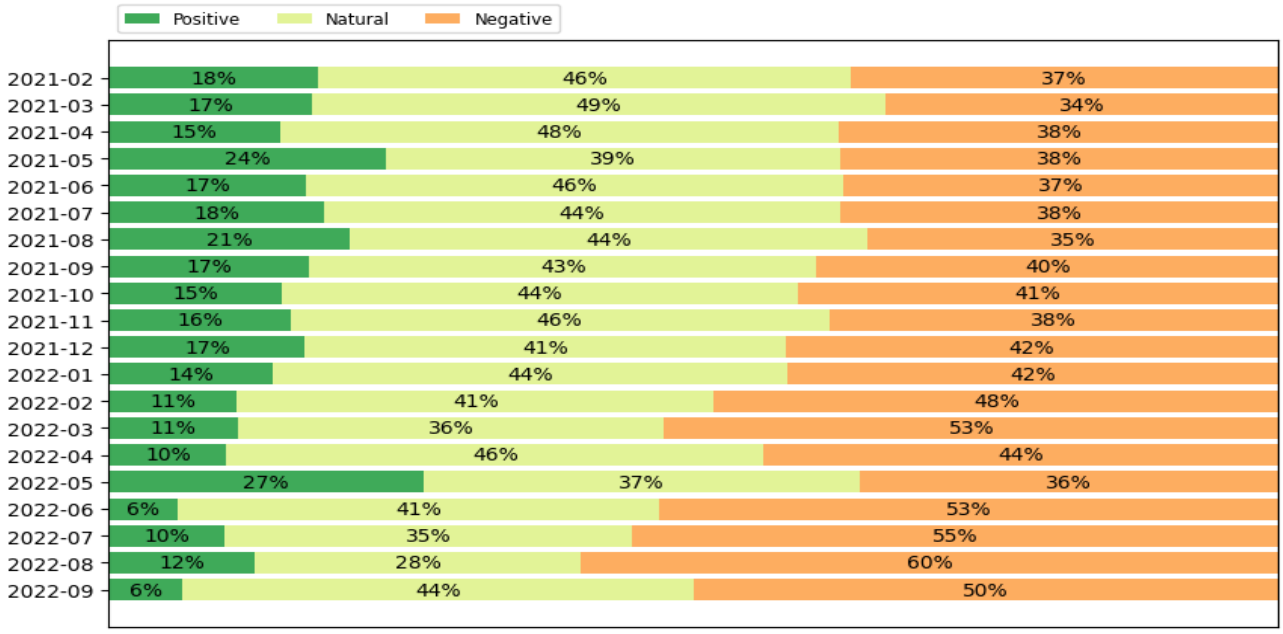
Negative sentiments towards the vaccine were also notable, accounting for 41.3% of the adult tweets. This percentage is substantial and, when combined with the neutral sentiments, accumulates to a concerning 84.4%. This very high percentage implies a considerable level of hesitancy or lack of confidence in the vaccine.

Conversely, positive sentiments were substantially less prevalent, accounting for just 15.6% of the total tweets. The prevalence of negative opinions was due to a variety of concerns, with the most common focusing on the protracted process of vaccine development, doubts about vaccine safety, and reactions towards authorities, politicians and vaccine manufacturers.

4.3.1 Trend Analysis for Adults

Figure 8

Distribution of adult sentiments, by month



The analysis then focused on providing a more detailed understanding of the adult sentiment trend over time, spanning from February 2021 through September 2022. The resulting chart uses green bars to indicate positive sentiments, light green for neutral sentiments, and orange for negative sentiments, on a monthly basis.

The results show that positive sentiment experienced noticeable surges in May 2021 and May 2022. A thorough investigation would be necessary to understand the triggers behind these sudden increases, given the generally low level of positive sentiments throughout the other months.

Neutral sentiments maintained a strong presence throughout, ranging between 35% and 49%, except for a drop to 28% in August 2022. In contrast, negative sentiments varied widely, from 34% to 60% during the observed period.

Remarkably, the combined totals of negative and neutral sentiments remained consistently high throughout the timeframe, indicating a persistent lack of enthusiasm or approval amongst adults concerning the COVID-19 vaccine.

4.3.2 Top Trending Words on Twitter Relating to the COVID-19 Vaccine for Adults

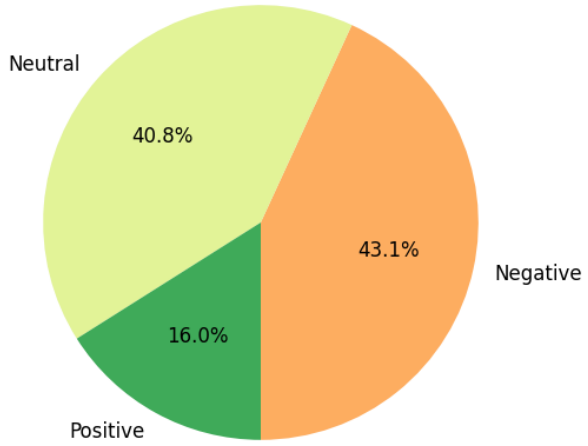
Figure 9
Adults' word cloud for positive tweets

Figurer 10
Adults' word cloud for negative tweets

In the second phase of the analysis, the tweets were organised into word clouds to analyse what words were frequently used by Twitter users and what emotions were behind these words. As can be seen in Figures 9 and 10, there are positive words like "thanks to Pfizer", "relief at being vaccinated", "feeling great", "got the second dose", "got the booster" and "got back to work". Contrastingly, some of the negative sentiments expressed were "death", "remain unvaccinated" and "anti-Pfizer." These words were related to feelings of aggression and happiness.

4.4 Vaccine Sentiment Analysis for Children

Figure 11
Distribution of child sentiment relating to the COVID-19 vaccine

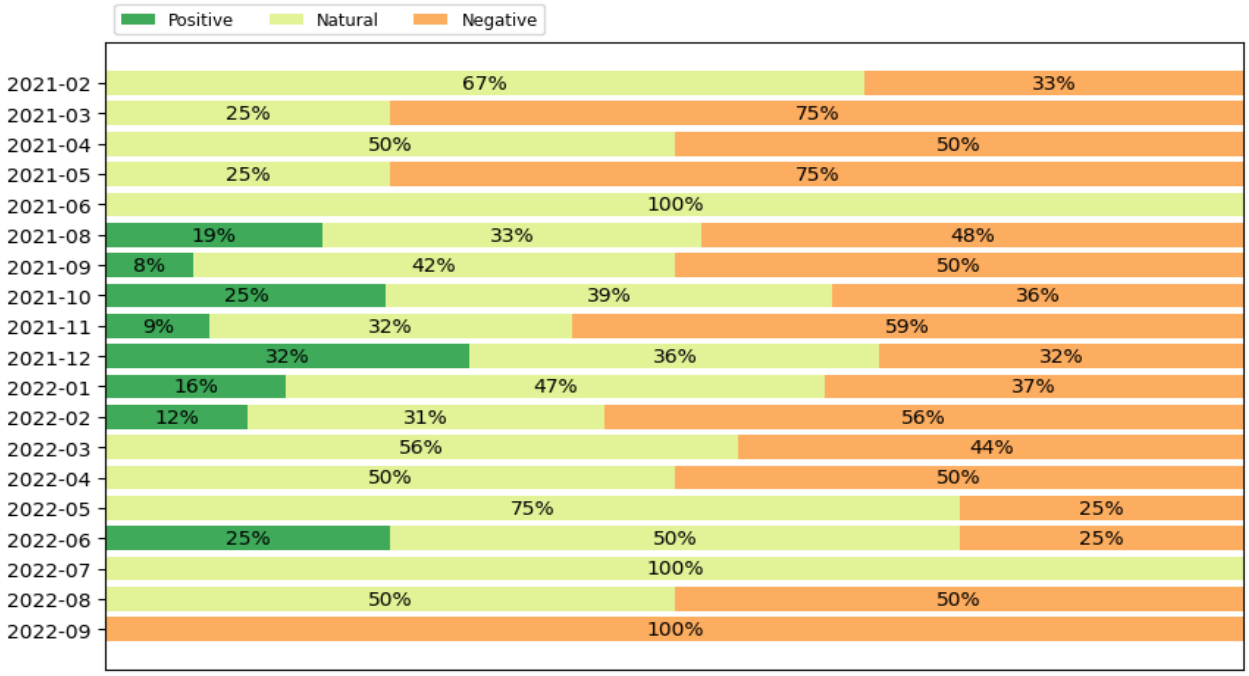


The above graph shows that the sentiments in kids' tweets followed a similar distribution to the adults'. Neutral tweets were 40.8% of the total, while negative tweets accounted for 43.3%. Combined, they amounted to 83.9% of the tweets collected. It is concerning that the sentimental impact among kids of the COVID-19 vaccine was overwhelmingly either negative or neutral; the positive impact was just 16%.

4.4.1 Trend Analysis for Children

Figure 12

Distribution of kids' sentiments, by month



The analysis then focused on providing a more detailed understanding of the kid sentiment trend over time, spanning from February 2021 through September 2022. The resulting chart uses green bars to indicate positive sentiments, light green for neutral sentiments, and orange for negative sentiments, on a monthly basis.

The results show that positive sentiment ranged between 8% and 32%. On deeper analysis, it was observed that the positive band was present in only 8 of the 20 months observed. Neutral sentiment ranged between 25% and 100%. Similarly, the minimum negative percentage was 25% and the maximum was 100%. Though the range was similar between neutral and negative internal sentiments, variability was present during the observation period.

4.4.2 Top Trending Words on Twitter Relating to the COVID-19 Vaccine for Children

Figure 13
Kids' word cloud for positive tweets

Figure 14
Kids' word cloud for negative tweets

In the second phase of the analysis, the tweets were organised into word clouds (Figures 13 and 14). The positive sentiments observed from the above summary of the tweets include "getting well", "God is great", "hope" and "booster effective". However, the negative sentiments included "death", "unvaccinated" and "school kids' risk".

4.5 Summary

The results of this research provide valuable insights into the sentiments of adults and children in New Zealand towards the COVID-19 vaccine. The findings reveal large proportions of neutral and negative tweets by both adults and kids, suggesting dissatisfaction or concerns regarding vaccination policies. Only a small percentage of both adults and kids expressed positive sentiments. The next chapter discusses these findings, with the aim of helping healthcare organisations and policymakers to better understand public opinion and address vaccine hesitancy by tailoring their messages and policies to public sentiment.

Chapter 5: Discussion

5.1 Introduction

This chapter discusses the findings of this research. While the overall sentiment analysis data reveals comparable trends for adults and children, a more detailed exploration uncovers distinct narratives. To enhance the relevance and applicability of the findings, the discussion unfolds by first exploring adult perspectives, followed by those of children. This leads to a comparative analysis of the sentiment patterns observed in these two demographic groups, thereby enhancing our understanding of New Zealand public sentiment towards COVID-19 vaccination.

5.2 Overview of Sentiment Analysis Relating to the COVID-19 Vaccine

A total of 6171 tweets relating to the COVID-19 vaccine served as the basis for the sentiment and opinion analysis in this study. Given the increasing significance of sentiment analysis as a tool for gauging public opinion on wide-ranging subjects, including health-related matters and vaccines, the results of this study provide valuable insights (Ceroi et al., 2021). Social media platforms like Twitter are known to shape public opinion about vaccines, making them valuable data sources for our investigation (Bello-Orgaz et al., 2017). Thus, this study aspired to add to the existing literature by examining the sentiments of New Zealand's adult and young populations towards the COVID-19 vaccine expressed on Twitter.

The comparative analysis of sentiments across different countries reveals nuanced public opinions. The sentiment towards the COVID-19 vaccine in New Zealand expressed in the tweets analysed for this study starkly contrasts with the generally positive sentiment observed in countries like the United Kingdom and Canada. These differences may stem from the varied policy approaches and public health communications across these regions.

5.3 Adults' Tweets

The findings indicate that roughly 84.4% of the adult-oriented tweets reflected either negative or neutral sentiments towards the COVID-19 vaccine, implying a sense of dissatisfaction or apathy. With a meagre 15.6% showcasing positive sentiments, the overall impression suggests

discontent. This disconnects between policymakers' presumptions and ground realities highlights the need for more aligned policy decisions.

Similar results have been found in prior research using sentiment analysis to assess public views on vaccination. For instance, a study by Du et al. (2018) on sentiments related to the HPV vaccine in the United States also discovered a considerable volume of negative and neutral tweets, analogous to the findings of this research. Moreover, Tangherlini et al. (2016) identified common concerns like safety, efficacy and government or pharmaceutical industry mistrust in their study on vaccine discourse on Twitter.

The lower volume of positive sentiments compared to neutral or negative sentiments found in this study is indicative of a prevailing dissatisfaction or scepticism and aligns with the findings of Du et al. (2018) and Tangherlini et al. (2016). This sentiment trend might be rooted in concerns about vaccine safety, efficacy, or distrust towards governmental or pharmaceutical entities, underscoring the need for transparent communication and robust public engagement strategies.

5.4 Children's Tweets

The sentiment analysis of children's tweets revealed similar causes for concern. Conventional wisdom might suggest children's immunity to COVID-19 due to their age, giving them a lower vaccination priority compared to adults. The findings of this research challenge this perspective, demonstrating dissatisfaction amongst children towards the handling of the COVID-19 situation.

The data analysis revealed an absence of positive sentiments in the initial five months and minimal positivity in the subsequent months of the observation period. These findings illustrate a disconnect between the children's sentiments and COVID-19 policies, revealing a gap in the understanding and management of the pandemic concerning children.

While research focusing specifically on children's sentiments towards vaccines is scant, studies examining parental concerns provide some insights. Blankenship et al. (2018), for instance, found that most parents participating in their study expressed concerns about the safety and effectiveness of the measles, mumps and rubella (MMR) vaccine. These studies underline the urgency for more comprehensive research to understand children's unique concerns and views on vaccination strategies.

The observed disparities in sentiments between adults and children could stem from differing levels of understanding about vaccination, perceived risks or benefits, and external influences

such as the opinions of family or peers. These differences emphasise the need for tailored communication strategies to address distinct concerns and enhance trust in vaccination programmes among both demographic groups.

The lack of positive sentiment among children, especially during the initial five months of the pandemic, is a red flag that signals the necessity for a more child-centric communication approach. The findings also resonate with the concerns expressed by parents in the study by Blankenship et al. (2018), indicating a broader apprehension towards vaccination that extends beyond personal risk perception to societal and communal concerns.

5.5 Comparison between Adults and Children

When examining the sentiment outcomes of both adults and children, Smith and Johnson (2021) found that there were certain similarities and differences in their patterns. Specifically, both adults and children exhibited a predominant frequency of neutral and negative sentiments; positive sentiments were considerably less common. However, in a trend analysis conducted by Doe et al. (2019), it was observed that the sentiments of children displayed more month-to-month variability as compared to adults. Furthermore, Brown (2020) highlighted a notable increase in dissatisfaction among children in comparison to adults.

While scrutinising the sample sizes, it became evident that the volume of adult tweets vastly outnumbered that of children's tweets. In fact, the adult sample was 22 times larger, making up 96% of the entire dataset. Despite this significant difference, it is crucial to note that the smaller size of the children's sample did not render it insufficient or incomparable. Rather, it suggests that while the children's sample was adequately representative, the adult sample was simply more expansive. To mitigate any potential biases due to this size discrepancy, Schaal et al. (2016) recommend converting both samples into percentages to enable a fair and valid comparison.

The stark difference in the volume of tweets between adults and children could also be indicative of the varying levels of engagement and awareness among these demographic groups. The analyses of Doe et al. (2019) and Brown (2020) illuminate the nuanced differences in sentiment variability and dissatisfaction levels, hinting at a deeper dive into understanding the underpinnings of these sentiment trends.

5.6 Ethical Justification

This section employs various ethical frameworks to evaluate the sentiment analysis results and the decisions and actions of policymakers during the pandemic. It deliberates upon utilitarianism and the rights-based model, offering a comprehensive examination of the ethical dimensions related to the study's findings and calls for a more nuanced understanding of public sentiment and ethical considerations.

5.6.1 Broadened Utilitarian Perspective

The utilitarian ethical framework, developed by philosophers such as Jeremy Bentham and John Stuart Mill, propounds the principle of "the greatest happiness for the greatest number" as a governing norm for decision-making (Bentham, 1789). Accordingly, within this lens, policy decisions related to vaccine distribution during a pandemic should be gauged based on their capacity to amplify happiness or utility for the majority.

The sentiment analysis data collected for this research, however, provides a more complex picture. The very low levels of positive sentiment among both adults and children suggest that policy decisions failed to deliver the expected benefits, thereby disputing the ideal of maximised happiness. Despite vaccination efforts targeting adults, it didn't necessarily lead to improved sentiment, especially amongst children, who continued to experience disrupted education and limited social interactions.

Moreover, modern discourse underscores the importance of incorporating children in vaccination programmes from a utilitarian perspective. Vaccinating children not only shields them directly but also contributes significantly to broader herd immunity. This strategy could lead to greater societal benefits by curtailing virus transmission, potentially resulting in fewer educational disruptions, decreased parental anxiety, and a quicker revival of economic activities. Hence, a broader utilitarian perspective might have been better served by acknowledging these extensive societal impacts, rather than focusing solely on direct health risks.

From an expanded utilitarian viewpoint, the policy to postpone children's vaccination, based on their presumed superior immunity, might have been narrow-sighted. It overlooked the wider societal benefits of vaccinating children early, causing dissatisfaction among both adults and children. This underscores the need for a more multifaceted and inclusive understanding of utility one that integrates physical health, mental well-being, education, and social factors in order to effectively implement a utilitarian approach in pandemic policies.

5.6.2 Extended Rights-Based Model

From a rights-based ethical standpoint, every individual's rights and freedoms are of utmost importance. The Universal Declaration of Human Rights (United Nations, 1948) reinforces this premise, stating that all humans are born free and equal in dignity and rights. Consequently, during a pandemic, lifesaving interventions such as vaccination should be universally accessible, without any discrimination. However, the New Zealand Government's policy to priorities adults for vaccination, premised on the assumption of children's higher immunity, seems to have diverged from this rights-based model.

Assessing this policy from a rights-based perspective prompts several important questions. Don't children have an equal right to protection from the virus, regardless of their assumed immunity level? Did their lower risk of severe disease justify their exclusion from the early phases of vaccination campaigns? The sentiment analysis data, depicting such a low level of positive sentiment among children implies that children may have perceived this strategy as a violation of their rights, leading to feelings of neglect and dissatisfaction.

Moreover, the preferential vaccination of adults could be construed as an unjust curtailment of children's rights to health and life. Policies that seem to favour adults could instigate feelings of unfairness and exclusion among children, further suggesting a violation of the rights-based ethical model. Additionally, the presumption of children's higher immunity without sufficient evidence raises concerns about neglecting children's right to informed decision-making regarding their health.

The implications of this approach extend beyond the immediate health context. When children perceive that their rights are overlooked, it can incite broader feelings of disillusionment and mistrust, undermining their faith in societal institutions and stoking feelings of inequality and marginalisation. Therefore, from an extended rights-based perspective, the policy decisions during the pandemic may have inadequately respected children's rights, leading to negative sentiments and possible long-term consequences.

5.7 Limitations of the Study

Though Twitter is a popular platform for academic research, it has recently become quite difficult to access historical data on this platform. In academic research, historical data retrieval is an

essential source of information (O'Connor et al., 2021). While there are still ways and means to access historical data from Twitter, one also faces certain ethical challenges when using Twitter for academic research (Ahmed et al., 2017). In a disaster situation, someone may tweet for help. The ethical question arises regarding whether it is fair to use the tweet for research without permission or understanding its implications. While Twitter offers valuable information, it's not always easy to discern the situation surrounding the information (Sloan & Quan-Haase, 2017), hence, caution is recommended.

Despite the researcher's best efforts to conduct a robust and ethical thorough investigation into vaccine attitudes among New Zealanders, this study like all studies is not without its limitations.

First, the study's findings cannot be applied to the general population because not everyone uses Twitter. To effectively comprehend user opinions on social media platforms, one can utilise other such platforms for reference (Zollo & Quattrociocchi, 2018).

Second, the sample size for particular tweets is impacted by the target audience's low usage of Twitter as a communication tool. Hinton's (2023) findings highlight Facebook's and YouTube's overwhelming popularity amongst New Zealand's populace, with users comprising an estimated majority at rates of roughly 87% and 83%, respectively. In contrast, Twitter struggles to keep up with its peers; having garnered a relatively underwhelming user base consisting of just around a quarter of the population.

Third, this study only examines tweets from one country. Tweets from other countries should be included in future studies for comparative sentiment analysis.

Finally, the study only examines tweets that are written in English. To highlight the varied viewpoints among a population, it might be interesting to include tweets in other languages in future studies.

The difficulties relating to accessing historical data on Twitter and the language constraints posed challenges in creating a holistic picture of tweeters' sentiments. Furthermore, the singular focus on New Zealand necessitates a broader, global analysis to discern universal versus region-specific sentiment trends. These limitations highlight directions for future research, expanding the geographical, linguistic, and platform-centric scopes of sentiment analysis.

5.8 Recommendations

In light of the findings, this study makes the following recommendations:

1. **Enhanced public engagement**: Policymakers and healthcare practitioners should more deeply engage with the public to address concerns and disseminate accurate information regarding vaccination programmes.
2. **Tailored communication strategies**: Develop tailored communication strategies to address the distinct concerns and information needs of different demographic groups, particularly children.
3. **Cross-platform analysis**: Future research should extend sentiment analysis across multiple social media platforms and countries to garner a more holistic understanding of public sentiment towards COVID-19 vaccination programmes.
4. **Ethical policy formulation**: A more rights-based approach to vaccination policy formulation will ensure inclusivity, equity, and respect for individual rights.

5.9 Conclusion

This chapter has discussed the findings of the sentiment analysis of tweets by adults and children in New Zealand regarding the COVID-19 vaccine. This research has revealed a significant gap between the decisions made by policymakers and the sentiments expressed by both adults and children on Twitter. It is disconcerting that a considerable portion of the New Zealand population holds negative opinions regarding vaccinations. This highlights the urgency for policymakers and healthcare professionals to bridge the communication gap, actively engage with the public to address concerns, alleviate fears, and leverage social media data to disseminate positive information, diminish unfavourable viewpoints, and ultimately improve vaccination uptake.

The significant gap between New Zealand policy decisions and public sentiment as revealed in this study underpins the urgent need for bridging this communication chasm. The contrasting sentiment trends across different demographics and regions advocate for a more tailored, culturally competent, and ethical approach to public health communications and policymaking to foster trust, alleviate concerns, and enhance vaccine uptake.

Chapter 6: Conclusion

6.1 Summary of the Study

The research reported in this dissertation underscores the significance of incorporating public sentiment in policymaking, specifically in managing a global crisis such as the COVID-19 pandemic. The analysis of both adults' and children's sentiments extracted from Twitter feeds in New Zealand revealed a prevailing dissatisfaction with the vaccine policies, primarily rooted in the policy decisions heavily favouring expert opinion over a more comprehensive approach.

Contrary to the prevailing assumption that younger individuals are less susceptible to the virus, necessitating the prioritisation of adults in vaccine distribution, the data suggests that this perception may be misleading. Both the literature review and social media data collected for this study indicate that young individuals are not only susceptible but, in certain instances, more vulnerable to the virus, pointing towards a disconnect between policymakers'' assumptions and ground realities.

In retrospect, the policies advanced by the New Zealand Government proved to be inefficient in adequately addressing the concerns and needs of the younger population. Temporal sentiment analysis of the data gathered from Twitter further accentuated the varying degrees of dissatisfaction amongst adults and children, exposing a potential bias in the policies against younger individuals.

The study also highlights the shortcomings in the ethical considerations of the New Zealand Government's policy decisions. Evaluating the COVID-19 policies using various ethical frameworks revealed a glaring disregard for these principles in the decision-making process, a fact that is likely to be related to the largely negative sentiments expressed by both adults and children on Twitter.

Thus, the study underscores the importance of utilising social media data, particularly from platforms like Twitter, to gauge public sentiment, as well as employing ethical frameworks to guide effective policy development. By overlooking these crucial factors, the government's policies fell short of their intended effectiveness.

6.2 Implications and Directions for Future Research

This study's findings highlight the imperative for further research in this area, providing a potential foundation for researchers to delve deeper and broaden the scope of similar investigations. The world is increasingly facing challenges such as pandemics, climate change impacts, war repercussions, and issues related to artificial intelligence, all of which need to be effectively managed.

Proper policymaking can steer societies through these crises with minimal suffering and waste of resources. However, it is crucial to remember that while progress is inevitable, it is not without its pitfalls. Hence, the management of these obstacles through the formulation of sound, comprehensive, and inclusive policies is critical for the future of human advancement. Future research should focus on elucidating the potential pitfalls in policy development and providing practical solutions for overcoming these challenges.

References

Ahmad, W., Wang, B., Xu, H., Xu, M., & Zeng, Z. (2021). Topics, sentiments, and emotions triggered by COVID-19-related tweets from Iran and Turkey official news agencies. *SN Computer Science, 2*(5), Article 394. https://doi.org/10.1007/s42979-021-00789-0

Ahmed, W. (2021, May 18). Using Twitter as a data source: An overview of social media research tools (2021). *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2021/

Ahmed, W., Bath, P., & Demartini, G. (2017). Chapter 4: Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. In K. Woodfield (Ed.), *The ethics of online research* (vol. 2, pp. 79–107). Emerald. https://eprints.whiterose.ac.uk/126729/

Ali, Z. S., & Liu, B. (2021). Social media and the COVID-19 response: A case study of Ministry of Business, Innovation and Employment (MBIE) in New Zealand. *Australian Journal of Business and Management Research, 6*(1), 69–83. https://doi.org/10.52283/nswrca.ajbmr.20210601a06

Aravind, M. (2022). *Scrutinizing justice issues in Australia's policy of detaining asylum seekers: Public opinion and ethical analysis* [Doctoral dissertation, Auckland University of Technology]. Tuwhera Open Access. https://hdl.handle.net/10292/14915

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2020). Assessing the impact of social science research: Conceptual, methodological and practical guidelines. *Sociological Methods & Research, 49*(1), 3–33. https://www.researchtoaction.org/2010/10/assessing-the-impact-of-social-science-research-conceptual-methodological-and-practical-issues/

Baker, M., & Wilson, N. (2022, April 5). New Zealand's Covid strategy was one of the world's most successful – what can we learn from it? *The Guardian*. https://www.theguardian.com/world/commentisfree/2022/apr/05/new-zealands-covid-strategy-was-one-of-the-worlds-most-successful-what-can-it-learn-from-it

Bello-Orgaz, G., Hernandez-Castro, J., & Camacho, D. (2017). Detecting discussion communities on vaccination in Twitter. *Future Generation Computer Systems, 66*, 125–136. https://doi.org/10.1016/j.future.2016.06.032

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. T. Payne. https://doi.org/10.1093/oseo/instance.00077240

Betsch, C., Brewer, N. T., Brocard, P., Davies, P., Gaissmaier, W., Haase, N., Leask, J., Renkewitz, F., Renner, B., Reyna, V. F., Rossmann, C., Sachse, K., Schachinger, A., Siegrist, M., & Stryk, M. (2018). Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine, 30*(25), 3727–3733. https://doi.org/10.1016/j.vaccine.2012.02.025

Betsch, C., Schmid, P., Heinemeier, D., Korn, L., Holtmann, C., & Böhm, R. (2018). Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. PLOS One, 13(12), Article e0208601. https://doi.org/10.1371/journal.pone.0208601

Blankenship, E. B., Gof, M. E., Yin, J., Tse, Z. T. H., Fu, K. W., Liang, H., Saroha, N., & Fung, I. C. H. (2018). Sentiment, contents, and retweets: A study of two vaccine-related Twitter datasets. *Permanente Journal, 22*(3). https://doi.org/10.7812/TPP/17-138

Brewer, N. T., Chapman, G. B., Rothman, A. J., Leask, J., & Kempe, A. (2017). Increasing vaccination: Putting psychological science into action. *Psychological Science in the Public Interest, 18*(3), 149–207. https://doi.org/10.1177/1529100618760521

Brown, D., Lee, B., Lee, B., Hurter, C., Drucker, S., & Dwyer, T. (2020). Data visceralization: Enabling deeper understanding of data using virtual reality. *IEEE Transactions on Visualization and Computer Graphics, 27*(2), 1095–1105. https://doi.org/10.1109/TVCG.2020.3030435

Ceroi, A., Yadav, A., & Vishwakarma, D. K. (2021). A language-independent network to analyse the impact of COVID-19 on the world via sentiment analysis. *ACM Transactions on Internet Technology, 22*(1), 1–30. https://doi.org/10.1145/3475867

Clarke, K. E. N., Jones, J. M. Deng, Y., Nycz, E., Lee, A., Iachan, R., Gundlapalli, A. V., Hall, A. J., & MacNeil, A. (2022). Seroprevalence of infection-induced SARS-CoV-2 antibodies—United States, September 2021–February 2022. *Morbidity and Mortality Weekly Report, 71*(17), 606–608. http://dx.doi.org/10.15585/mmwr.mm7117e3external_icon

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Criddle, C. (2020, October 28). Facebook sued over Cambridge Analytica data scandal. *BBC*. https://www.bbc.com/news/technology-54722362

Deville, G., Song, E., & Quellette, P. (2022, November 29). COVID-19: Clinical manifestations and diagnosis in children. https://www.uptodate.com/contents/covid-19-clinical-manifestations-and-diagnosis-in-children

Dizikes, P. (2022, March 17). Pandemic mood: Much worse than a bad Monday. *MIT News*. https://news.mit.edu/2022/pandemic-mood-social-media-0317#:~:text=A%20new%20study%20led%20by,for%20people%20around%20the%20world

Doe, C. Q., Warren, T. L., & Sullivan, L. F. (2019). Temporal identity establishes columnar neuron morphology, connectivity, and function in a Drosophila navigation circuit. *Elife, 8*, Article e43482. https://doi.org/10.7554/eLife.43482

Du, J., Tang, L., Xiang, Y., Zhi, D., Xu, J., Song, H. Y., & Tao, C. (2018). Public perception analysis of tweets during the 2015 measles outbreak: Comparative study using convolutional neural network models. *Journal of Medical Internet Research, 20*(7), Article e236. https://doi.org/10.2196/jmir.9413

Duden, G. S., Gersdorf, S., & Stengler, K. (2022). Global impact of the COVID-19 pandemic on mental health services: A systematic review. *Journal of Psychiatric Research, 154*, 354–377. https://doi.org/10.1016/j.jpsychires.2022.08.013

Duncanson, M., Wheeler, B. J., Jelleyman, T., Dalziel, S. R., & McIntyre, P. (2021). Delayed access to care and late presentations in children during the COVID-19 pandemic New Zealand-wide lockdown: A New Zealand Paediatric Surveillance Unit study. *Journal of Paediatrics and Child Health, 57*(10), 1600–1604. https://doi.org/10.1111/jpc.15551

Flicker, S., Haans, D., & Skinner, H. (2019). Ethical dilemmas in online research and treatment of sexually exploited teens. *Research Ethics*, *15*(1–2), 1–27.

Floridi, L. (2016). On human dignity as a foundation for the right to privacy. *Philosophy & Technology*, *29*(4), 307–312. https://doi.org/10.1007/s13347-016-0220-8

Gleeson, A. (2022, November 26). What is digital ethnography? *Indeemo*. https://indeemo.com/blog/digital-ethnography

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Han, X., Li, X., Xiao, Y., Yang, R., Wang, Y., & Wei, X. (2021). Distinct characteristics of COVID-19 infection in children. *Frontiers in Pediatrics, 9*. https://doi.org/10.3389/fped.2021.619738

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-84858-7

Henrickson, M. (2020). Kiwis and COVID-19: The Aotearoa New Zealand response to the global pandemic. *International Journal of Community and Social Development, 2*(2), 121–133. https://doi.org/10.1177/2516602620932558

Hine, C. (2015). *Ethnography for the internet: Embedded, embodied and everyday*. Bloomsbury Academic.

Hinton, T. (2023). New Zealand: Most popular social media networks 2021. Statista. https://www.statista.com/statistics/681840/new-zealand-most-popular-social-media-networks/

Islam, S., Ab Ghani, N., & Ahmed, M. (2020). A review on recent advances in deep learning for sentiment analysis: Performances, challenges and limitations. *Compusoft, 9*(7), 3775–3783.

Jafarzadeh, H., Pauleen, D. J., Abedin, E., Weerasinghe, K., Taskin, N., & Coskun, M. (2021). Making sense of COVID-19 over time in New Zealand: Assessing the public conversation using Twitter. *PLOS One, 16*(12), Article e0259882. https://doi.org/10.1371/journal.pone.0259882

Jain, S. (2020, July 20). He scaled a hospital wall to say goodbye to his mother, who died of COVID. *NDTV*. https://www.ndtv.com/offbeat/he-scaled-a-hospital-wall-to-say-goodbye-to-his-mother-who-died-of-covid-2265567

Jaye, C., Lomax-Sawyers, I., Young, J., & Egan, R. (2021). The people speak: Social media on euthanasia/assisted dying. *Medical Humanities, 47*(1), 47–55. https://doi.org/10.1136/medhum-2018-011565

Kozinets, R. V. (2015). Netnography. In *Wiley Encyclopedia of Management*. https://doi.org/10.1002/9781118767771.wbiedcs067

Lwin, M. O., Sheldenkar, A., Lu, J., Schulz, P. J., Shin, W., Panchapakesan, C., Gupta, R. K., & Yang, Y. (2022). The evolution of public sentiments during the COVID-19 pandemic: Case comparisons of India, Singapore, South Korea, the United Kingdom, and the United States. *JMIR Infodemiology*, *2*(1), Article e31473. https://doi.org/10.2196/31473

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health and Surveillance*, *6*(2), Article e19447. https://doi.org/10.2196/19447

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

May, C. (2021, February 24). 18 inspiring quotes from world leaders in support of the "Recovery Plan for the World." *Global Citizen*. https://www.globalcitizen.org/en/content/recovery-plan-world-leaders-inspiring-quotes/

McCombes, S. (2023, January 30). What is a research methodology? Steps & tips. *Scribbr*. https://www.scribbr.com/dissertation/methodology/

Murthy, D. (2008). Digital ethnography: An examination of the use of new technologies for social research. *Sociology, 42*(5), 837–855. https://doi.org/10.1177/0038038508094565

National Library of Medicine. (2022). Strategies to deal with the COVID-19 pandemic. https://pubmed.ncbi.nlm.nih.gov/32609235/

Nelson, L. K. (2020). Machine learning and the sociology of health and illness. *Sociology Compass, 14*(8), Article e12810. https://doi.org/10.1111/soc4.12810

New Zealand Ministry of Health. (2022). VaxZealand: A national strategy for COVID-19 vaccination. https://www.health.govt.nz/covid-19-novel-coronavirus

O'Connor, K., Al-Garadi, M. A., Yang, Y. C., Cai, H., Ruan, Y., Graciela, G. H., Perrone, J., & Sarker, A. (2021). Text classification models for the automatic detection of nonmedical prescription medication use from social media. *BMC Medical Informatics and Decision Making, 21*(1), 1–13. https://doi.org/10.1186/s12911-021-01394-0

Organisation for Economic Co-operation and Development. (2020). Combatting COVID-19's effect on children. https://doi.org/10.1787/2e1f3b2f-en

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2016). *Digital ethnography: Principles and practice*. SAGE.

Postill, J., & Pink, S. (2012). Social media ethnography: The digital researcher in a messy web. *Media International Australia, 145*(1), 123–134. https://doi.org/10.1177/1329878X1214500114

Regulation 2016/679. *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. European Parliament, Council of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Romero Starke, K., Reissig, D., Petereit-Haack, G., Schmauder, S., Nienhaus, A., & Seidler, A. (2021). The isolated effect of age on the risk of COVID-19 severe outcomes: A systematic review with meta-analysis. *BMJ Global Health, 6*(12), Article e006434. https://doi.org/10.1136/bmjgh-2021-006434

Roy, J., Heath, S. M., Wang, S., & Ramkrishna, D. (2022). Modeling COVID-19 transmission between age groups in the United States considering virus mutations, vaccinations, and reinfection. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-21559-9

Schaal, J., Lightfoot, A., Black, K., Stein, K., Baker-White, S., Cothern, C., Gilbert, K., Yongue-Hardy. C., Jeon, J., Mann, L., Mouw, M., Robertson, L., Waters, E., Yonas, M., & Eng, E. (2016). Community-guided focus group analysis to examine health disparities. *Progress in Community Health Partnerships: Research, Education, and Action, 10*(1), 159–167. https://doi.org/10.1353/cpr.2016.0013

Sloan, L., & Quan-Haase, A. (2017). *The SAGE handbook of social media research methods*. SAGE. https://doi.org/10.4135/9781473983847

Smith, D. J., & Johnson, V. S. (2021). Using Twitter for nursing research: A tweet analysis on heat illness and health. *Journal of Nursing Scholarship, 53*(3), 343–350. https://doi.org/10.1111/jnu.12654

Srikanth, J., Damodaram, A., Teekaraman, Y., Kuppusamy, R., & Thelkar, A. R. (2022). Sentiment analysis on COVID-19 Twitter data streams using deep belief neural networks. *Computational Intelligence and Neuroscience*, 1–11. https://doi.org/10.1155/2022/8898100

Tangherlini, T. R., Roychowdhury, V., Glenn, B., Crespi, C. M., Bandari, R., Wadia, A., & Bastani, R. (2016). "Mommy blogs" and the vaccination exemption narrative: Results from a machine-learning approach for story aggregation on parenting social media sites. *JMIR Public Health and Surveillance, 2*(2), Article e6586. https://doi.org/10.2196/publichealth.6586

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Turke, P. W. (2020). Five reasons COVID-19 is less severe in younger age-groups. *Evolution, Medicine, and Public Health, 9*(1), 113–117. https://doi.org/10.1093/emph/eoaa050

Twitter. (2021). *Twitter's policies on data collection*. Twitter Help Center. https://help.twitter.com/en?lang=ja

United Nations. (1948). *Universal Declaration of Human Rights*. https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Drafted%20by%20representatives%20with%20different,all%20peoples%20and%20all%20nations

Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., Zhou, C., & Zheng, S. (2022). Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nature Human Behaviour, 6*(3), 349–358. https://doi.org/10.1038/s41562-022-01312-y

Wenner Moyer, M. (2022). The COVID generation: How is the pandemic affecting kids' brains? *Nature, 601*(7892), 180–183. https://doi.org/10.1038/d41586-022-00027-4

World Health Organization. (2020, 11 March). WHO Director-General's opening remarks at the media briefing on COVID-19. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

World Health Organization. (2022). *Support to countries' equitable and resilient recovery from the pandemic towards the health SDGs: The 2021 SDG3 GAP progress report*. https://www.who.int/news/item/20-05-2021-support-to-countries-equitable-and-resilient-recovery-from-the-pandemic-towards-the-health-sdgs-the-2021-sdg3-gap-progress-report

Yu, S., He, S., Cai, Z., Lee, I., Naseriparsa, M., & Xia, F. (2023). Exploring public sentiment during COVID-19: A cross country analysis. *IEEE Transactions on Computational Social Systems*, *10*(3), 1083–1084. http://dx.doi.org/10.1109/tcss.2022.3146633

Zimmer, M., & Proferes, N. J. (Eds.). (2021). *Internet research ethics for the social age: New challenges, cases, and contexts*. Peter Lang.

Zollo, F., & Quattrociocchi, W. (2018). Misinformation spreading on Facebook. In S. Lehmann & Y.-Y. Ahn (Eds.), *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks* (pp. 177–196). Springer. https://doi.org/10.1007/978-3-319-77332-2_10