Thesis Dissertation

ANALYSIS OF TWITTER SENTIMENT IN RESPONSE TO GLOBAL PANDEMIC EVENTS: A CASE STUDY OF COVID-19

Nicolas Theodosiou

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

May 2024

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

Analysis of Twitter Sentiment in Response to Global Pandemic Events: A Case Study of COVID-19

Nicolas Theodosiou

Supervisor

Dr. George Pallis

Thesis submitted in partial fulfilment of the requirements for the award of degree of Bachelor in Computer Science at University of Cyprus

May 2024

Acknowledgments

First and foremost, I would like to extend my deepest gratitude to my supervisor, Dr. George Pallis, for his invaluable support and guidance throughout the course of this research. His expertise and insightful feedback have been crucial in shaping both the direction and the outcome of this work.

I am also immensely grateful to Dr. Dimosthenis Stefanidis for providing the datasets necessary for this study, and for his guidance throughout the project. His expertise in data analysis and his willingness to share his knowledge have greatly enhanced the quality of my research.

Special thanks go to my classmate, Simona Uteva, with whom I collaborated closely on this topic. Our joint efforts and mutual support helped us overcome the challenges we faced during our research. Her perspective and input have significantly contributed to the success of this project.

Lastly, I wish to thank my friends and family for their unwavering support and encouragement throughout my academic pursuits. Their belief in my capabilities and their constant encouragement have kept me motivated and focused, even during the most challenging times.

Abstract

This thesis examines the evolution of public sentiment on Twitter during the COVID-19 pandemic. By analyzing tweets from the onset of the pandemic in January 2020 to February 2023, this study identifies significant shifts in public sentiment across different phases of the pandemic, related to varying COVID-19 restrictions, vaccine rollouts, and other pivotal events.

Utilizing a dataset encompassing millions of tweets from users worldwide, this research applies advanced sentiment analysis techniques to categorize tweets into emotional responses such as fear, joy, sadness, and anger. The analysis is enriched by correlating these sentiments with specific pandemic-related events and measures, such as lockdowns and vaccination milestones. Additionally, the study conducts a comparative analysis across key countries, such as Great Britain and India, to uncover regional differences in public sentiment.

Key findings indicate a direct correlation between government-imposed restrictions and significant swings in public emotion, particularly in expressions of negative sentiments. Positive sentiments also experienced fluctuations corresponding to the progress in COVID-19 vaccine developments and deployment. Furthermore, the thesis explores the evolution of topics over time, revealing how public focus shifted from immediate health concerns to long-term societal and economic impacts.

The insights derived from this study contribute to our understanding of social media's role in shaping public sentiment during a global crisis, offering implications for policymakers, health authorities, and communication strategists in managing public emotion and information dissemination during ongoing and future global challenges.

Contents

1. Introduction

1.1 Topic Overview

In the era of digital communication, social media platforms such as Twitter have transcended their roles as mere networking sites to become powerful tools for real-time sentiment analysis, especially during global crises [1]. The COVID-19 pandemic, which began in early 2020, has been one of the most significant global events of the 21st century, transforming communication and interaction patterns worldwide. As countries grappled with unprecedented public health challenges, Twitter emerged as a crucial platform for millions to express opinions, fears, and hopes, reflecting broad public sentiment in real-time [2]. This thesis explores how Twitter data can be analyzed to understand shifts in public sentiment throughout the pandemic, offering insights into how global crises influence public emotions and discourse.

Sentiment analysis, the computational study of people's opinions, sentiments, and emotions expressed in text, has important applications in fields ranging from marketing to political science. Despite its potential, sentiment analysis is accompanied by many challenges, including the rapid evolution of language, the ambiguity of expressions, and contextual nuances. The COVID-19 pandemic presents a unique case study for sentiment analysis due to its global impact and the large amount of discussions it generated on platforms like Twitter [3]. Traditional research often fails to capture the complex, evolving nature of public sentiment during such prolonged crises, focusing instead on static or localized snapshots. This research aims to bridge this gap by utilizing a continuous and in-depth sentiment analysis of Twitter feeds, thus contributing to a deeper understanding of dynamic public sentiment.

The literature on sentiment analysis within social media contexts has extensively documented the utility of these platforms for understanding public opinion, particularly during crises. However, gaps remain in the existing research, notably in longitudinal studies that track sentiment over extended periods and during globally impactful events like the COVID-19 pandemic. Studies such as those by Smith et al. (2020) have shown that Twitter can effectively reflect shifts in public mood in response to immediate events, but they often do not capture prolonged sentiment trends or the delayed effects of cumulative events [4]. Furthermore, while many researchers have utilized sentiment analysis to gauge public health messaging effectiveness, few have integrated these findings with policy changes and global news events to draw comprehensive conclusions about public sentiment trajectories. This thesis builds on the foundation laid by previous work, aiming to fill these gaps by providing a detailed analysis of both the immediate and extended emotional responses of the public to the pandemic, as seen through Twitter.

1.2 Problem Statement

The COVID-19 pandemic has caused unprecedented global changes, affecting various aspects of human life and eliciting a wide range of emotional responses that have manifested particularly on social media platforms like Twitter. This surge in digital expressions presents a challenge in understanding how public sentiment has evolved in response to the pandemic's unfolding events and varying governmental actions. Addressing this challenge, the study focuses on deciphering the shifts in emotions such as fear, joy, sadness, and anger, particularly in reaction to significant developments like lockdowns and vaccine rollouts. The primary problem this research seeks to solve is the difficulty in tracking and interpreting these rapid changes in public mood, which are critical for policymakers and health communicators.

1.3 Research Questions

This study is motivated by several key research questions designed to dissect and understand the complex dynamics of public sentiment on Twitter during the COVID-19 pandemic:

How has public sentiment evolved in response to the progression of the COVID-19 pandemic? This question seeks to map the shifts in global public sentiment over time, linking these changes directly to the unfolding of the pandemic. By analyzing these emotional responses against the pandemic timeline, the study aims to uncover the patterns and triggers of sentiment changes.

What is the relationship between key pandemic-related events and shifts in public sentiment on Twitter? This question examines the influence of significant events, such as public health announcements and policy changes, on public sentiment. Focusing on correlating these events with subsequent sentiment fluctuations provides insights into the public's reactive dynamics on social media.

How does sentiment compare across different countries and regions? This comparative analysis focuses on understanding the geographical variations in sentiment, highlighting how different cultural and governmental contexts influence public emotional responses.

How do specific topics discussed on Twitter contribute to emotional fluctuations? This question involves conducting topic modeling to categorize tweets and analyze how discussions within specific thematic areas fluctuate in emotional tone and intensity.

These research questions aim to provide a comprehensive understanding of the emotional landscape on Twitter during the COVID-19 pandemic.

2. Related Work

2.1 Previous Research

The sentiment analysis of Twitter data during the COVID-19 pandemic has provided invaluable insights into public opinion and reactions to unfolding events. Researchers have utilized various methodologies to explore how sentiments and discourse themes have evolved in response to the pandemic's developments. These studies provide an insight into the collective psyche, highlighting concerns, misinformation trends, and shifts in public sentiment over time.

One significant contribution in this area was made by Joanne Chen Lyu and colleagues [5], who pioneered a methodical approach by utilizing machine learning algorithms to dissect a substantial amount of tweets gathered during the early months of the pandemic. Their methodology extended beyond basic sentiment analysis to include topic modeling, which enabled the identification and categorization of main themes within the discussions.

By applying tools such as Latent Dirichlet Allocation (LDA) and sentiment classifiers trained on nuanced language models, their research delineated how topics like public health policies, economic impact, and community responses evolved alongside growing public awareness and changing pandemic dynamics. This detailed analysis not only spotlighted dominant emotional reactions but also traced how these sentiments shifted from fear and uncertainty to cautious optimism as the pandemic progressed.

Another related research effort by Brown and Zhou [6] explored the dynamism of public sentiment in response to specific pandemic-related events using a combination of time-series analysis and event study methodologies. Their work correlated the timings of significant announcements, like lockdowns and vaccine progress, with shifts in sentiment polarity captured through Twitter data.

By leveraging regression analysis techniques alongside sentiment scoring tools such as VADER (Valence Aware Dictionary and sEntiment Reasoner), they quantified the immediate and sustained emotional impacts of these events. They found that public sentiment was significantly affected by such events, with noticeable shifts towards negativity during lockdown periods and a surge in positivity following positive vaccine news.

In another crucial study, Waters et al. [7] offered a unique perspective by focusing on the sentiment among entrepreneurs on Twitter, utilizing network analysis to map how sentiments and misinformation spread across different user clusters. They employed sophisticated data mining techniques to identify key influencers and their role in disseminating information or misinformation.

By analyzing retweet networks and the propagation patterns of tweets, this research highlights the role of influential Twitter accounts in shaping discourse, demonstrating how sentiment and misinformation propagation can be mapped and potentially mitigated through targeted communication strategies.

These studies underscore the many different approaches to analyzing Twitter data, each contributing distinct insights into public sentiment dynamics during the pandemic. The methodologies employed provide a deep dive into:

- **Machine Learning and Topic Modeling:** As seen in Lyu's work, these techniques are crucial for parsing large datasets to extract thematic and emotional nuances, enabling researchers to track the evolution of public discourse over time.
- **Event-Driven Analysis and Time-Series Techniques**: Utilized by Brown and Zhou, these methods help in understanding the temporal

impact of specific events on public sentiment, crucial for policymakers aiming to gauge and manage public reactions.

- **Network Analysis:** Applied by Waters and colleagues, this approach reveals the interconnections within social networks, offering insights into how information and emotions flow through communities, which is essential for designing targeted communication strategies.

Together, these studies underscore the diverse approaches to analyzing Twitter data, each contributing unique insights into the ways in which public sentiment about the pandemic has been shaped by events and information flow.

2.2 Gap in Literature

Existing research on Twitter sentiment analysis during the COVID-19 pandemic has shed light on the public's emotional responses, utilizing diverse methodologies to quantify sentiment changes over time. However, there remain several underexplored areas and methodological gaps that could be addressed in future studies to enhance the depth and applicability of sentiment analysis:

Integration of Multimodal Data Sources

Most current studies focus solely on textual data from Twitter, potentially overlooking the rich context provided by multimodal content such as images, videos, and emojis used in tweets. These elements play a crucial role in conveying sentiments and could provide a more nuanced understanding of public emotions. Future research could incorporate these multimodal aspects to gain a comprehensive view of sentiment and its expression on social media.

Cross-Platform Sentiment Analysis

While Twitter is a valuable source for capturing real-time public sentiment, it is just one of many social media platforms where users express their opinions and emotions. There is a gap in studies comparing sentiments across different platforms like Facebook, Instagram, or Reddit, which might reveal platform-specific discourse patterns and broader, more diverse public sentiments.

Longitudinal Studies Beyond the Pandemic

The majority of sentiment analysis studies provide snapshots limited to specific phases of the pandemic. There is a need for longitudinal studies that track sentiment changes over extended periods, potentially before, during, and after the pandemic, to understand how sentiments stabilize or evolve as societies recover and adapt to new normal.

Advanced Analytical Techniques for Subtle Sentiments

Current tools and methodologies may lack the sensitivity to detect subtle sentiments or mixed emotions, which are especially prevalent in complex crisis situations like a pandemic. There is a gap in developing and applying more advanced analytical techniques that can decipher these subtleties in sentiment more effectively.

Socio-Demographic Analysis of Sentiment

Few studies delve into how different demographic groups express sentiments differently on social media. Understanding these differences can provide insights into how various segments of the population react to public health crises, allowing for more targeted and effective communication strategies.

3. Background

Delving into the foundational concepts, this chapter outlines the crucial role of social media analytics and sentiment analysis in understanding public sentiment and behavior on platforms such as Twitter. It explains how sentiment on social media is shaped by global events, emphasizing the relevance of Twitter as a significant source of data. The chapter also details the technical tools and programming languages used in capturing and analyzing this data.

3.1 Conceptual Foundations

3.1.1 Sentiment Analysis

Sentiment analysis is a critical tool for data scientists and researchers aimed at understanding public perceptions and emotions as expressed through social media. This technique is important for converting the vast amounts of unstructured text available on platforms like Twitter into structured data that can be analyzed to discern public sentiment trends.

It operates through the application of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from source materials [8]. This process is crucial for understanding the population's sentiment during crises such as the COVID-19 pandemic.

This analysis includes a variety of sentiment analysis techniques, ranging from simple polarity checks, which categorize text as positive or negative, to more complex emotional assessments that can detect sadness, joy, and context-based sentiment.

3.1.2 Social Media Analytics

Social media analytics is a key component of contemporary data analysis, leveraging data from platforms like Twitter, Facebook, and Instagram to get insights into user behavior, preferences, and societal trends.

This analysis involves more than just observing visible metrics such as likes and retweets; it also aims to understand the patterns of engagement and interaction among users, and how content spreads across networks. The ability to analyze these interactions provides valuable insights into public mood and opinion trends, which can be useful for researchers studying social dynamics [9].

Moreover, social media analytics also aids in the identification of influential users and key opinion leaders who play crucial roles in information dissemination. The methodologies used, such as network analysis and machine learning, help parse through large datasets to identify trends, anomalies, and patterns that are not immediately obvious.

3.2 Technical Background

3.2.1 Twitter as a Data Source

As one of the most influential platforms where users express opinions, share news, and engage in public discourse, Twitter offers a rich repository of real-time data that reflects current social dynamics and public sentiments.

Twitter's data is particularly valuable because it is both large in volume and granular, providing tweets that can be analyzed for content, sentiment, and user engagement metrics such as retweets, likes, and replies [3]. This makes it an ideal platform for studying the spread of information, the evolution of public opinion, and the dynamics of social interactions on a global scale.

The structure of Twitter data, which includes metadata such as timestamps, geographic information, and user demographics, allows researchers to conduct detailed analyses that can track changes in sentiment over time, correlate events with shifts in public mood [4], or map how information spreads geographically. The accessibility of Twitter's API further makes the extraction of this data easy, enabling researchers to collect vast amounts of information systematically.

3.2.2 Data Analysis Tools

This section emphasizes the critical role of the tools used for managing and interpreting the vast amounts of data gathered from social media platforms like Twitter.

Key among the tools utilized was JupyterLab, an interactive development environment that helps with code development, data visualization, and collaborative analysis. JupyterLab was useful in allowing for real-time data manipulation and visualization, with its ability to integrate text, code, and visual outputs into a single document streamlining the workflow [10].

To manage the large volumes of data efficiently, the research also leveraged the computational power of the Computer Science Department's Cluster. This high-performance computing cluster was crucial for executing resource-intensive tasks that are beyond the capabilities of ordinary personal computers. It supported large-scale data processing tasks, ensuring that the analysis was conducted with high efficiency and accuracy [11].

The combination of JupyterLab for detailed, interactive data analysis and the processing capabilities of the department's cluster significantly enhanced the management and analysis of large datasets. This synergy made the exploration of complex data more efficient and effective, underscoring the value of integrating advanced tools in research methodologies.

3.2.3 Software and Programming Languages

Exploring the software and programming languages essential for executing the methodologies of this study, a particular focus is placed on Python and R, two of the most widely used programming languages in data science and statistical analysis, due to their powerful libraries and frameworks for handling large-scale data analysis and machine learning tasks.

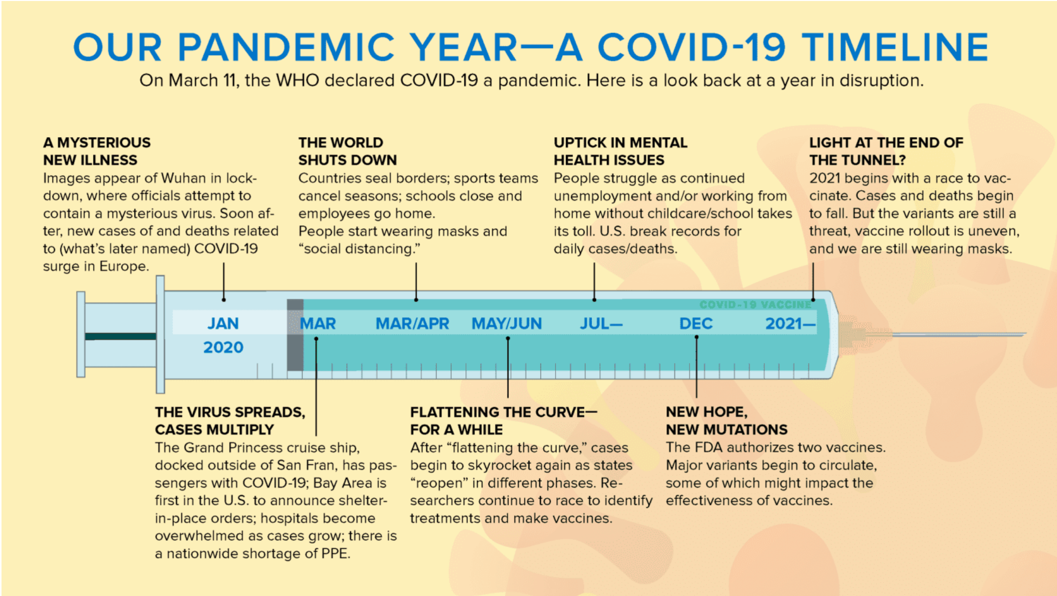
Python is renowned for its versatility and strength in data manipulation and analysis, supported by libraries such as Pandas, NumPy, and Scikit-learn. These tools are essential for data cleaning, transformation, statistical modeling, and machine learning, making Python very useful for the complex sentiment analysis conducted in this thesis [12].

R is another critical tool, valued for its statistical computing capabilities. It was utilized particularly for its sophisticated environment, which is highly optimized for statistical analysis, aiding in the analysis of large datasets [13].

The combination of Python's data handling capacities and R's statistical strengths was crucial for effectively processing and analyzing the Twitter data collected.

3.3 Overview of the Pandemic

The COVID-19 pandemic, initially identified in Wuhan, China in December 2019, the novel coronavirus, SARS-CoV-2, rapidly escalated into a global health crisis, declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [14].



As the virus spread across continents, countries implemented varying levels of public health responses. Initial reactions included travel bans, quarantines, and the closure of public spaces. As we can see in the image [57], by April 2020, over half of the world's population was under some form of lockdown [15], an unprecedented public health measure aimed at curbing the virus's spread. These lockdowns led to significant disruptions in daily life, economic activities, and a palpable shift in public behavior and sentiment.

The economic impact of the pandemic was profound, with the global economy entering a recession [16]. Unemployment rates soared as businesses closed or reduced operations. Governments worldwide launched fiscal stimulus packages to mitigate the economic downturn and support unemployed workers and affected industries.

On the social front, the pandemic dramatically altered everyday life. Education systems shifted to online learning, and teleworking became the norm for many, leading to a reevaluation of work-life balance and digital connectivity's role in maintaining societal functions. Social isolation and the stress of the pandemic caused mental health issues, bringing attention to the need for mental health support systems [17].

Scientifically, the pandemic spurred a global race to develop vaccines and treatments. By the end of 2020, multiple vaccines had been developed and approved for emergency use [18], marking a significant milestone in the fight against the virus. Vaccination campaigns varied by country, with disparities in vaccine availability and acceptance creating challenges in achieving widespread immunity.

Understanding the impact of the COVID-19 pandemic is vital for analyzing the corresponding shifts in public sentiment and behavior observed on Twitter. The detailed timeline of events and responses provides a critical context for exploring how public discourse on the platform evolved in response to the various phases of the pandemic.

4. Methodology

This chapter outlines the methodologies and tools used to collect, analyze, and interpret the data crucial for evaluating the impact of the COVID-19 pandemic on public sentiment via Twitter. It includes detailed descriptions of the data collection process, applications of natural language processing for sentiment analysis, and the statistical methodologies used to detect changes in emotional expressions.

4.1 Datasets

4.1.1 Tweet Emotions & Characteristics

The emotional content and interactive features of tweets provide an insight into user behavior and sentiment on social media. This subsection dives into the datasets that track a range of emotions such as hate, joy, and stress, alongside key tweet characteristics like favorites, hashtags, and retweets.

4.1.1.1 Data Sources

The analysis conducted on Tweet emotions and characteristics relies on data gathered from two primary sources: Crunchbase and the Twitter API. Each source was selected based on its relevance to the research objectives, which include analyzing the geographic distribution and sentiment of Twitter users. This section details the methodology behind the acquisition and processing of data from these sources, outlining how each contributes to the comprehensive dataset used for analysis.

Crunchbase

Crunchbase is a comprehensive database of companies and startups worldwide, often used in academic and market research to gather insights about business ecosystems.

- **Data Extraction:** Data was extracted specifically for users listed in the Crunchbase database who have linked Twitter accounts. The focus was primarily on their Twitter user IDs and associated geographic locations, including country and state.
- **Methodology:** An automated script was developed to query the Crunchbase database through its API, extracting relevant user data.
- **Purpose:** This data served as a foundational layer to identify and correlate the geographic distribution of users and their tweeting patterns, facilitating a demographic analysis in subsequent stages.

Twitter API

The Twitter API provides programmatic access to Twitter data, allowing for the collection of tweets and associated metadata.

- **Data Collection:** Tweets were collected over a period from January 2018 to February 2023, capturing a broad timeline to analyze trends over time.
- **Volume:** A total of 67,457,434 tweets were gathered from 107,613 users across 59 countries, providing a substantial dataset for statistical analysis.
- **Methodology:** Custom scripts were used to interface with the Twitter API, utilizing both historical access and real-time streaming where necessary to compile the dataset.
- **Filtering and Selection Criteria:** Tweets were filtered by user ID as sourced from Crunchbase to ensure data consistency and relevance to the user base identified. Further filtering based on language (only tweets in English) was applied to refine the dataset for analysis.
- **Purpose:** This extensive collection of tweets forms the primary dataset for sentiment analysis and temporal trend evaluation in the thesis. It allows for the examination of changes in public sentiment over time, assessment of the impact of global events on Twitter discourse, and comparison of geographic sentiment distribution.

4.1.1.2 Methodology and Tools

This section describes the methodologies and analytical tools employed to conduct emotional and sentiment analysis within the scope of this research. To ensure a comprehensive understanding of the sentiments shared on Twitter, a variety of state-of-the-art tools and models have been used. Each tool has been selected for its specific ability to detect and classify emotional content, sentiment, hate speech, offensive content, and stress indicators among Twitter users.

Joy, Optimism, Anger and Sadness Detection

- **Tool:** Hugging Face's RoBERTa-base for Emotion [19]
- **Output:** Probability score between 0-1 indicating likelihood of the specific emotion
- **Methodological Reference:** Barbieri et al. (2020) [20] present TweetEval, a unified benchmark and comparative evaluation framework for tweet classification, focusing on various linguistic properties, including emotional content. This model leverages the robust training capabilities of RoBERTa, finely tuned on large-scale social media text data, to capture nuanced emotional expressions effectively.

Fear Detection

- **Tool:** Hugging Face's RoBERTa-base for Fear Emotion [21]
- **Output:** Probability score between 0-1 indicating likelihood of fear
- **Methodological Reference:** Utilizes the approach described by Liu et al. (2019) [22], focusing on the "RoBERTa: A Robustly Optimized BERT Pretraining Approach." This model optimizes BERT architectures for more dynamic and context-aware understanding, crucial for identifying expressions of fear among Twitter users.

Sentiment Analysis

- **Tool:** AllenNLP Models for Sentiment Analysis [23]
- **Output:** Binary classification of sentiment (Positive: 1, Negative: -1)
- **Methodological Reference:** Follows the robust optimization principles of RoBERTa as discussed by Liu et al. (2019) [22], which enhances the model's ability to discern fine-grained sentiment from textual data.

Hate Speech Detection

- **Tool:** Hugging Face's RoBERTa-base for Hate Speech [24]
- **Output:** Probability score between 0-1 indicating likelihood of hate speech
- **Methodological Reference:** Utilizes frameworks developed by Barbieri et al. (2020) [20] that are specifically tuned to identify hate speech in social media posts, allowing for accurate and reliable moderation of content.

Offensive Content Detection

- **Tool:** Hugging Face's RoBERTa-base for Offensive Content [25]
- **Output:** Probability score between 0-1 indicating likelihood of offensive content
- **Methodological Reference:** Also based on Barbieri et al. (2020) [20], which provides a robust mechanism to classify tweets for offensive content, aiding in comprehensive content analysis.

Stress Detection

- **Tool:** MedDL for Stress Classification [26]
- **Output:** Binary classification (Stressed: 1, Not Stressed: 0)
- **Methodological Reference:** Described in Scepanovic et al. (2020) [27], this tool extracts medical entities from social media, applying deep learning to identify stress indicators effectively, which is crucial for mental health studies.

4.1.1.3 Dataset Overview

In order to conduct a thorough analysis of Twitter interactions and sentiment related to the COVID-19 pandemic, multiple datasets were used, each designed to capture distinct aspects of Twitter user behavior and engagement. These datasets range from basic Twitter metrics, which provide an insight to user interactions, to more complex engagement metrics and emotional analyses that offer insights into user sentiment and behavior. This section outlines the core attributes and specific utilities of each dataset.

Dataset 1: Basic Twitter Metrics

- **Core Attributes:** TwitterUserID, TweetID, Created At
- **Engagement Metrics:** Retweet Number, Favorites Number, Mentions Number, Hashtags Number, URLs Number
- **Data Utility**: This dataset serves as a foundational layer for analyzing user engagement, providing a macroscopic view of Twitter interactions.

Dataset 2: Enhanced Twitter Engagement Metrics

- **Core Attributes:** TwitterUserID, TweetID, Created At
- **Engagement Metrics:** Retweet Number, Favorites Number, Mentions Number, Hashtags Number, URLs Number
- **Sentiment Analysis Scores:** Stress Score, Offensive Score, Hate Score
- **Data Utility:** The inclusion of content-specific scores such as Stress, Offensive, and Hate scores allows for a nuanced analysis of the nature and tone of discourse, giving a deeper insight into the sentiment and behavioral patterns on Twitter.

Dataset 3: Emotional Analysis

- **Core Attributes:** TwitterUserID, TweetID, Created At

- **Engagement Metrics:** Retweet Number, Favorites Number, Mentions Number, Hashtags Number, URLs Number
- **Sentiment Analysis Scores:** Joy, Sadness, Anger, Fear, Optimism, Sentiment
- **Data Utility:** This dataset provides an emotional breakdown of tweets, offering valuable data points for assessing the emotional responses of users during different phases of the COVID-19 pandemic.

Dataset 4: User Details

- **Core Attributes:** TwitterUserID
- **User Information:** Country, State, City, Gender
- **Data Utility:** This dataset focuses on the demographic attributes of Twitter users, providing a demographic layer to the engagement and emotional data. By correlating user demographics with engagement metrics and emotional expressions, this dataset helps in understanding how different demographic groups react to and engage with pandemic-related content on Twitter.

4.1.2 Tweet Topic

This chapter is dedicated to methodology, model training, and performance details of the developed topic classifier model for analysis of Twitter data during the COVID-19 pandemic. This model puts tweets in different thematic categories, providing a categorized dataset on which further sentiment analysis can be carried out.

4.1.2.1 Model Description

The Topic Classifier is a fine-tuned version model based on the roberta-base model [22], trained on a dataset of 256,000 news headlines from New York

Times articles spanning between the year 2000 and 2023. This model was chosen because of its good performance with text classification tasks. The fine-tuning process adapted this model to the specifics of the Twitter data, aligning the output categories to the most prevalent topics in COVID-19 related discussions on Twitter.

4.1.2.2 Methodology and Tools

To ensure the effectiveness and reliability of the model in classifying tweet topics accurately, a setup of hyperparameters and a structured training process were implemented. This section outlines the specific configurations used during model training, including learning rates, batch sizes, optimizer settings, and scheduler types.

Hyperparameters and Training Process

The following hyperparameters were used during the model training:

- Learning Rate: 5e-05
- Train Batch Size: 8
- Evaluation Batch Size: 8
- Optimizer: Adam with betas=(0.9, 0.999) and epsilon=1e-08
- LR Scheduler Type: Linear
- LR Scheduler Warmup Steps: 500
- Number of Epochs: 5

The model was trained over 5 epochs, recording every detail of the training and validation loss metrics for each epoch. In this regard, it is to be mentioned that the model showed a high gain of accuracy and other associated metrics over epochs, telling us about the effective learning of the model and adaptation to the training data.

Training Performance Metrics

The training results showcased progressive improvement in model performance:

- Epoch 1 recorded an accuracy of 88.65%, with a slight increase in validation loss, suggesting initial model adaptation.
- By Epoch 3, the model achieved an accuracy of 90.94%, with reduced validation loss.
- The final epoch saw the highest accuracy at 91.37%, with consistent gains in precision and recall.

Model Performance on Test Data

The performance of the model on the test set (51,200 cases) was robust, achieving good performance across all metrics. This demonstrates the model's capability to generalize well on unseen data.

Detailed performance by topic was as follows:

Topic	Precision	Recall	F1	Support
Sports	0.97	0.98	0.97	6400
Arts, Culture, and Entertainment	0.94	0.95	0.94	6400
Business and Finance	0.85	0.84	0.84	6400
Health and Wellness	0.9	0.93	0.91	6400
Lifestyle and Fashion	0.95	0.95	0.95	6400
Science and Technology	0.89	0.83	0.86	6400
Politics	0.93	0.88	0.9	6400
Crime	0.85	0.93	0.89	6400

The demonstrated metrics highlight the model's performance across a wide range of topics, showcasing its effectiveness in accurately classifying tweets into thematic categories. The topic classifier described has proven to be a good instrument for analyzing the thematic composition of tweets, providing a solid basis for further sentiment analysis.

4.1.2.3 Dataset Overview

This research utilizes a dataset comprising of tweet IDs, topic labels, and corresponding topic scores. To maintain the accuracy in topic classification, the analysis was restricted to records with a topic score exceeding 0.9. This threshold ensures a high level of confidence in the relevance and precision of the topic assignments, which is crucial for conducting a reliable analysis.

Topic Breakdown

- **Sports:** Captures tweets related to various sports events, athletes, and sports news. This category helps analyze how major sporting events, like the Olympics or the FIFA World Cup, influence Twitter engagement and sentiment globally.
- **Arts, Culture, and Entertainment:** Encompasses a wide range of subjects from movie releases and celebrity news to discussions about literature and visual arts. This topic is essential for gauging public interest in entertainment and cultural developments and their impact on social interactions during different periods.
- **Business and Finance:** Includes tweets related to the stock market, corporate affairs, and significant economic events. Analyzing this topic assists in understanding how economic trends influence public discourse and sentiment, particularly during economic downturns or booms.
- **Health and Wellness:** This topic covers tweets concerning health tips, pandemic updates, and wellness advice. During the COVID-19 pandemic, this category was particularly active, offering insights into public concerns and misinformation regarding health crises.

- **Lifestyle and Fashion:** Focuses on trends in fashion and lifestyle choices. This area explores how cultural and societal norms are reflected through Twitter, providing a view into consumer behavior and seasonal trends.
- **Science and Technology:** Involves discussions about technological advancements, innovations, and scientific discoveries. Tweets under this topic are important for understanding how information about science and technology spreads through social media and influences public opinion.
- **Politics:** Encompasses tweets about political debates, elections, policy changes, and political figures. This category is crucial for assessing how political events drive online discussions and polarize opinions.
- **Crime:** Includes coverage of criminal activities, law enforcement actions, and public safety announcements. Analyzing this topic helps understand the public's response to crime and safety concerns, which can vary significantly across different regions and times.

4.1.3 Regional Covid Data

This section presents an in-depth overview of the regional COVID datasets, which are comprised by detailed records of COVID-19 metrics such as cases, deaths, and vaccination rates, alongside government-imposed restrictions including mask mandates, school closures, and income support measures across various countries. The datasets are important for uncovering correlations between the spread of the virus, public health responses, and the socio-economic repercussions on populations.

4.1.3.1 Data Sources

COVID-19 Data

- **Source**: Our World in Data (OWID) - COVID-19 Data Repository [28]

- **Description**: This dataset, maintained and updated by Our World in Data, provides comprehensive global statistics related to COVID-19, including daily updates on cases, deaths, testing rates, and vaccination figures. The data is aggregated from multiple international sources, ensuring wide coverage and reliability. This dataset is critical for analyzing the spread and impact of the pandemic across different regions and times.

COVID-19 Restrictions

- **Source**: Our World in Data (OWID) - COVID-19 Policy Responses [29]
  Description: The COVID-19 Government Response Tracker, compiled by Our World in Data, offers a detailed catalog of policy responses implemented by governments worldwide in reaction to the pandemic. This includes data on public health measures such as mask mandates and school closures, as well as economic support actions like income support payments and employment protection. The dataset categorizes these interventions, providing a chronological overview and a comparative analysis across various jurisdictions.

4.1.3.2 Dataset Overview

Building on the datasets introduced earlier, this section delves into the specifics of the epidemiological data and governmental response measures that are crucial for the analysis of the pandemic. It presents a structured overview of each dataset, detailing the attributes that capture the progression of the pandemic and the policies employed.

COVID-19 Data

This dataset provides a detailed record of the COVID-19 pandemic's progression across various countries. It captures both cumulative and new figures, offering insights into daily changes and long-term trends in case numbers and mortality rates. The inclusion of vaccination data allows for an analysis of the impact of vaccination rollouts on pandemic dynamics. Each entry is associated with a country code and date, ensuring that the data can be analyzed over time and across different regions.

Description of COVID-19 Data

Attribute	Description
Country Code	ISO country codes
Date	Date of data entry
Total Cases	Cumulative number of confirmed cases
New Cases	Daily new cases
Total Deaths	Cumulative number of confirmed deaths
New Deaths	Daily new deaths
Positive Rate	The ratio of positive tests to total tests
Total Vaccinations	Cumulative number of vaccine doses administered

COVID Restrictions

This dataset tracks the various public health and social measures implemented by governments in response to the COVID-19 pandemic. Each record includes a comprehensive set of interventions, such as income support, workplace and school closures, and the imposition of mask mandates. The data provides a country-specific timeline of when different measures were enforced or relaxed, allowing for comparative analysis of their effectiveness and the extent of their enforcement.

Description of COVID-19 Restrictions

Attribute	Description
Country Code	ISO country codes
Date	Date of data entry
Income Support	Direct financial payments to people who lose their jobs or businesses

Workspace Closures	Closure of non-essential workplaces
Facial Coverings	Requirements for face masks in public areas
Vaccine Availability	Availability of COVID-19 vaccines to the public
School Closures	Closure of schools and universities
Stay Home Requirements	Orders to stay at home unless absolutely necessary
Close Public Transport	Shutdowns of public transportation systems
International Travel Controls	Restrictions on international travel
Restriction Gatherings	Limits on the size of gatherings
Debt Relief	Financial relief for debtors during the pandemic
Vaccination Policy	Policies regarding the vaccination of the population
Containment Index	A composite measure of the intensity of the containment measures

4.2 Data Pre-processing

4.2.1 Cleaning up

Efficient data cleaning and preprocessing are important for ensuring the integrity and usability of the dataset used in this analysis. This section elaborates on the methods used to prepare the data for a thorough assessment of Twitter interactions during the COVID-19 pandemic.

Removal of Incomplete Records

The dataset initially contained a number of records with missing information in critical fields, such as TweetID and Date, which are essential for any temporal analysis. Tweets with null values in these columns were removed to prevent inaccuracies in trend analysis and sentiment assessment. This step was crucial in maintaining a dataset that could yield reliable information about the Twitter dynamics over the specified period.

Standardization of Data Format

Variations in data format were observed, with discrepancies mainly in the structure of the tweet entries. Some records contained additional, non-standard columns due to data extraction anomalies or changes in the data collection API over time. A script was written to identify and remove any tweets that did not conform to the predefined schema of expected columns. This normalization ensured that subsequent data handling processes, such as transformations and aggregations, would operate smoothly and without error.

Gender Mapping

The dataset featured a diverse range of gender identifiers, reflecting the varied ways individuals choose to express gender identity on social platforms. To consolidate these into analytically useful categories, a comprehensive gender mapping strategy was implemented. The decision to map specific genders to broader categories was guided by both the need to respect the complexity of gender identities and the practical constraints of statistical analysis [30]. The mapping was as follows:

- **Male-Identified:** 'male', 'ftm' (female-to-male), 'transgender_male', 'transgender_man' were all mapped to 'male'.
- **Female-Identified:** 'female', 'transgender_woman', 'transgender_female', 'mtf' (male-to-female) were all mapped to 'female'.
- **Non-Specified/Non-Provided:** Entries such as 'not_provided' and 'nan' (not a number) were categorized under 'not_provided'.
- **Other/Non-Binary Identities:** Terms like 'gender_fluid', 'non_binary', and other non-conforming identifiers were grouped under 'other'.

The completion of these cleaning steps resulted in a polished dataset free from unnecessary noise and inconsistencies. This refinement allows for more accurate statistical modeling and ensures that the analysis can more reliably capture true patterns and insights from the data.

4.2.2 Merging Datasets

In order to conduct a holistic analysis of Twitter interactions during the COVID-19 pandemic, it was important to create a unified dataset that includes all aspects of user engagement, sentiment analysis, emotional response, and demographic details. This section details the process of merging the four distinct datasets described previously into a single comprehensive dataset.

Preparing Datasets for Integration

Each dataset came with its own set of attributes, centered around the common keys of TwitterUserID and TweetID. The initial step involved ensuring data consistency across all datasets:

- **Standardization**: Converting all data formats to align with those of the central dataset (Dataset 1: Basic Twitter Metrics). This includes formatting dates, numeric values, and categorical data to ensure compatibility.
- **Cleaning**: Prior to merging, datasets underwent a cleaning process to remove any duplicates or irrelevant records that could skew analysis, as discussed in the data cleaning section.

Alignment of Data Entries

- **Key Matching**: The merging was based on TwitterUserID and TweetID, ensuring that data linked to the same tweet and user across different datasets was accurately consolidated.

- **Synchronization**: Time-stamped data (such as tweet creation times and dates) were synchronized to allow for time-series analyses across different datasets.

Consolidation into a Master Dataset

- **Horizontal Merging**: The datasets were combined horizontally, adding columns from the Enhanced Twitter Engagement Metrics, Emotional Analysis, and User Details datasets to the Basic Twitter Metrics.
- **Validation**: After merging, checks were performed to validate the integrity of the merged data. This included verifying that no data was lost or incorrectly merged during the process.

Final Structure of the Master Dataset

- **Core Twitter Metrics**: Includes basic and enhanced engagement metrics such as retweets, favorites, mentions, hashtags, and URLs.
- **Sentiment and Emotion Scores**: Integrates detailed sentiment scores and emotional analysis data, providing insights into the psychological state of the Twitter community during the pandemic.
- **Demographic Details**: Enriches the dataset with demographic information such as country, state, city, and gender, facilitating demographic-specific analyses.

Utility of the Master Dataset

- **Multi-dimensional Analysis**: Allows for the examination of interactions between different variables, such as how demographic factors influence engagement and sentiment.
- **Temporal and Geographic Trends**: Supports detailed time-series and geographic analyses to observe how sentiments and public discourse evolved over different phases of the pandemic.

- **Policy Impact Evaluation**: Enables an assessment of how different COVID-19 related policies affected public sentiment and behavior on social media.

To address distinct research questions and allow for targeted analyses, the master dataset was further used to create three more specialized datasets. Each subset was designed to explore different dimensions of the interaction between Twitter activity and the COVID-19 pandemic, focusing on emotional responses, policy impacts, and topic engagement.

COVID Data Dataset

- **Purpose**: This dataset was made to analyze the correlation between COVID-19 case dynamics such as vaccination rates and death counts, and the emotional tone of tweets.
- **Selection Criteria**: From the master dataset, all records containing both emotional scores (e.g., joy, sadness, anger) and relevant COVID-19 epidemiological data (e.g., total vaccinations, new deaths) were extracted.
- **Utility**: By combining emotional analysis with COVID-19 health data, this dataset enables a nuanced exploration of how public sentiment on Twitter correlates with the progression of the pandemic and public health milestones.

COVID Restrictions Dataset

- **Purpose:** Aimed at assessing the impact of governmental policies and restrictions on public sentiment, this dataset isolates the effect of interventions like mask mandates and school closures.
- **Selection Criteria:** Records that included emotional scores and data on COVID-19 policy measures (e.g., facial coverings, school closures) were extracted to form this dataset.
- **Utility:** It helps provide an understanding of how various public health policies influence collective emotional responses on social media,

giving insights into public receptiveness and reaction to policy changes.

Tweet Topic Dataset

- **Purpose:** This dataset focuses on the relationship between specific topics of discussion and the emotional responses they cause.
- **Selection Criteria:** Only the data points that included topic labels (e.g., sports, politics) and corresponding emotional scores were selected.
- **Utility:** By examining the interplay between tweet topics and emotional sentiments, this dataset allows for an analysis of how different subjects impact public mood and engagement levels, highlighting topics that trigger significant emotional responses during the pandemic.

Data Processing and Integration Steps

For each subset, the following steps were followed:

- **Data Extraction**: Using scripted queries, relevant fields were extracted based on the predefined criteria for each dataset.
- **Data Integrity Checks**: After extraction, data integrity checks were conducted to ensure that no essential information was lost or misaligned during the segmentation process.
- **Final Adjustments**: Necessary adjustments were made to ensure each dataset was optimized for the specific analyses planned, including recalibrating certain variables and reformatting data structures as needed.

4.2.3 Region specific Datasets

To understand the geographical distribution of Twitter activity related to the

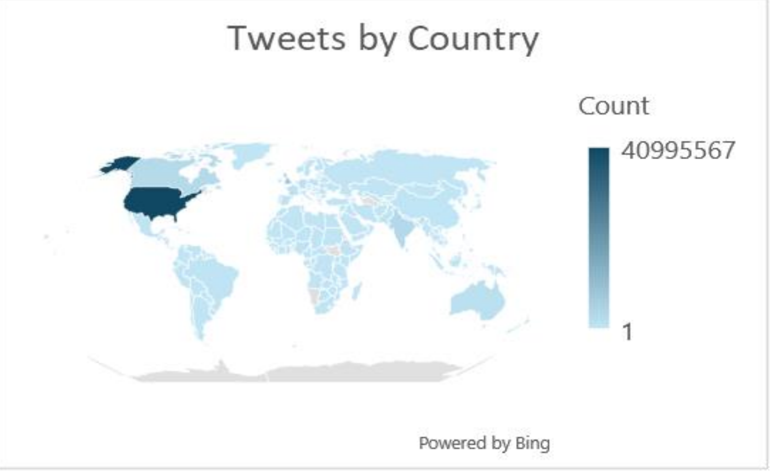
COVID-19 pandemic, an initial analysis was conducted on the dataset to identify the countries with the highest volume of tweets.

Identifying Top Tweeting Countries

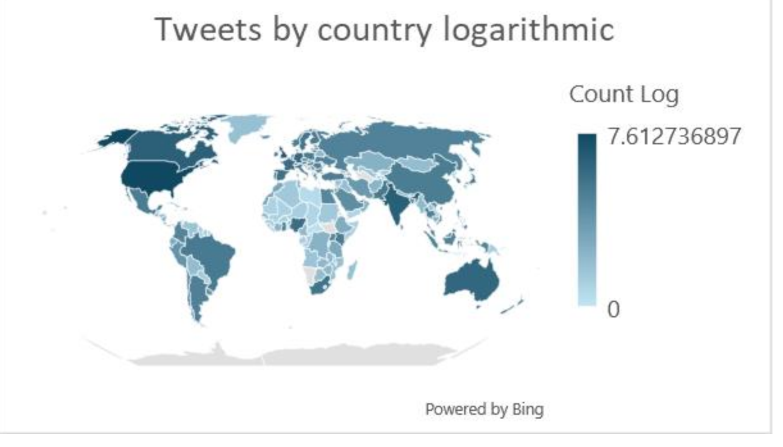
Rank	**Country**	**Tweet Count**
1	United States	40995567
2	United Kingdom	7433638
3	India	3154006
4	Canada	2685104
5	Australia	1279445
6	Germany	1185018
7	France	656305
8	Netherlands	580109
9	Nigeria	574320
10	Republic of Ireland	468000
11	Spain	448460
12	Singapore	438898
13	United Kingdom	431717
14	Switzerland	395376
15	Israel	371145
16	South Africa	315261
17	Sweden	301425
18	Italy	274802
19	United Arab Emirates	257793
20	Belgium	243459

Using the master dataset, a query was executed to aggregate the number of tweets per country. The results were summarized in a table listing the top 20 countries by tweet volume. This table serves as a fundamental reference for identifying regions with significant Twitter activity, which are critical to understanding the global response to the pandemic.

Visualization with World Heatmaps



First Heatmap (Raw Counts): A global heatmap was created to show the raw counts of tweets from each country. This visualization provides a straightforward depiction of Twitter activity distribution worldwide.



Second Heatmap (Logarithmic Scale): Given the disproportionate volume of tweets originating from the United States, a second heatmap using a logarithmic scale was developed. This approach adjusts for the wide range in tweet volumes, enhancing visibility for countries with fewer tweets and allowing for a more comparative analysis of global engagement.

These heatmaps are important in visualizing the extent and intensity of discussions related to COVID-19 across different regions, portraying discrepancies, and concentrations in social media engagement.

Country-Specific Datasets for Targeted Analysis

Following the identification of the top tweeting countries, specialized subsets of the master dataset were created for the three countries with the highest tweet volumes: the United States (USA), Great Britain (GBR), and India (IND). Each country-specific dataset was derived by filtering the master dataset for tweets originating from these countries. This process was replicated to create corresponding subsets from the COVID Data Dataset, COVID Restrictions Dataset and Tweet Topic Dataset.

The region-specific analysis and the creation of targeted datasets for the USA, GBR, and IND enable a focused examination of the pandemic's impact on public sentiment, policy responses, and prevalent topics within these regions. This study focuses on the analysis of the GBR and IND datasets, trying to draw insights about how different cultural, political, and social contexts influence public discourse and reactions to the pandemic.

4.3 Statistical Analysis

This section outlines the application of t-tests and fixed effects regressions to analyze emotional responses on Twitter, both globally and for specific regions (GBR, IND), before and after the onset of the COVID-19 pandemic. These statistical methods were selected to evaluate the immediate and long-term emotional impact caused by pandemic-related developments.

T-Tests

T-tests were employed to determine if there were statistically significant differences in the expressions of emotions before and after the onset of the COVID-19 pandemic [31]. This method was applied to:

- **Global Analysis**: Assess overall shifts in public sentiment on a worldwide scale.
- **Regional Analysis**: Examine specific changes within Great Britain, and India to understand regional nuances in emotional responses.
- **Application**: For each emotional category, t-tests compared the

average sentiment scores before the pandemic to those after its declaration.

- **Interpretation of Results**: The results highlight the initial emotional impact of the pandemic, indicating which emotions showed significant changes and potentially revealing varied regional reactions.

Fixed Effects Regressions

In this study, fixed effects regression models are utilized to examine the influence of the COVID-19 pandemic on key emotional responses captured in tweets. This statistical approach allows for an in-depth analysis of how public emotions such as fear and hate are affected by the pandemic, considering individual tweeting behaviors and external circumstances [32]. The model's details are as follows:

- **Dependent Variables**: The primary focus of the analysis involves several emotional metrics, which serve as dependent variables in our models. These include quantified measures of emotions such as fear, hate, and other relevant affective states extracted from the tweets.
- **Control Variables**: To ensure the reliability of our findings, the model includes tweet characteristics like likes and favorites as control variables. These allow us to adjust for the impact of tweet popularity and engagement on the emotional expressions being analyzed.
- **Instrumental Variable**: An instrumental variable (IV) is incorporated to distinctly identify the period of the COVID-19 pandemic. Tweets posted during this time are coded as 1, and those from before the pandemic as 0. This binary variable helps to isolate the effect of the pandemic from other temporal variations that could influence emotional expressions.
- **Grouping Criterion**: Tweets are grouped based on unique Twitter user IDs. This grouping enables the model to control for individual-specific effects, which are critical for distinguishing the changes in emotional responses due to the pandemic from each user's personality.

The regression analysis is structured to control for both tweet-specific characteristics and individual user effects, ensuring that the observed changes in emotions are specifically attributable to the impact of COVID-19. By maintaining user-specific constants and including controls and an instrumental variable, the models provide a detailed understanding of how the pandemic has influenced emotional responses on Twitter.

Integrating t-tests and fixed effects regressions enables a comprehensive understanding of how emotional responses to the pandemic were manifested initially and how they evolved over time. While t-tests provide snapshots of immediate changes at the pandemic's onset, fixed effects regressions track the ongoing emotional trajectories of individual Twitter users. This approach ensures a thorough examination of both immediate and sustained emotional responses both worldwide and across different geographical contexts.

4.4 Data Visualization

Data visualization plays a crucial role in this study, enabling the effective communication of complex data and insights. Various visualization techniques were employed to explore and represent the relationships between different variables, track changes over time, and highlight key trends in the data. This section describes the specific visualizations used to analyze correlations, changes, and trends within the gathered Twitter data related to COVID-19.

Heatmaps for Correlation Analysis

- **Emotion Correlations:** A heatmap was utilized to visualize the correlations between different emotions expressed in tweets. This approach helps to identify which emotions commonly co-occur, indicating underlying patterns in public sentiment.
- **COVID Data Correlations:** Another heatmap was created to examine the correlations between various COVID-19 data metrics, such as

case counts, death rates, and vaccination numbers. This visualization aids in understanding how these metrics interrelate and potentially influence public discourse.

- **COVID Restrictions Correlations:** A third heatmap focused on the relationships between different COVID-19 restrictions, like mask mandates and school closures. This visualization is key to discerning patterns and strengths of associations between various policy measures.

Bar Charts for Monthly Change Analysis

Bar charts were employed to analyze the percentage monthly change of each emotion relative to the mean emotion level, with green bars above the x-axis signifying a positive change and a red bar under the y-axis signifying a negative change. This method provides a clear and straightforward representation of how emotional expressions fluctuate over time in response to the evolving pandemic situation.

Combined Bar and Line Charts for Detailed Trend Analysis

- **Emotion and COVID Data Trends:** A combination of bar charts and line charts was used to depict the monthly change in specific emotions alongside corresponding changes in selected COVID-19 data points, such as new cases or vaccination rates. This dual-chart approach allows for a comparative analysis of how specific pandemic developments affect public emotions.
- **Emotion and Restriction Trends:** Similarly, this visualization technique was also applied to compare changes in emotions with the implementation or lifting of specific COVID-19 restrictions, providing insights into the emotional impact of policy decisions.

Line Charts for Topic and Emotion Dynamics

Line charts were used to track the monthly changes in the volume of tweets pertaining to specific topics, such as politics, health, or economy. Additionally, line charts were also used to show how the emotional content related to each topic has shifted over time, offering a dynamic view of the interplay between topic prevalence and associated sentiment.

5. Results

This chapter presents the findings of the statistical analyses and data visualizations performed on the compiled Twitter datasets. It includes a detailed examination of the emotional, topical, and regional dynamics captured in tweets during the COVID-19 pandemic. The results highlight significant emotional changes, the impact of specific COVID-19 policies, and the most frequent topics of discussion across different stages of the pandemic. The following subsections are structured to first address global trends and insights, followed by a focused analysis on the specific regions of Great Britain, and India. The findings provide an understanding of how public sentiment on Twitter has evolved in response to ongoing global events and interventions.

5.1 Worldwide Analysis

The worldwide analysis section focuses on the global patterns and trends of Twitter activity during the COVID-19 pandemic. This comprehensive analysis draws on the aggregated data from the master dataset, focusing on the changes in public emotions, the effectiveness of health policies, and shifts in topical discussions over time. Key findings are presented in the form of heatmaps, bar charts, and line charts to visualize the scale and scope of Twitter interactions across various countries. This section aims to uncover how global events linked to the pandemic have influenced sentiment and behavior on a global scale, offering insights into the impact of COVID-19 on social media.

5.1.1 Emotion Analysis

Exploring the shifts in public sentiment on Twitter during the COVID-19 pandemic, this analysis focuses on the emotions expressed across the platform. Utilizing the data derived from our master dataset, the analysis focuses on quantifying and interpreting the shifts in emotions such as joy, sadness, anger, and fear among global Twitter populations. Through a combination of statistical tests and correlation analyses, this section seeks to uncover the emotional responses to pandemic-related developments and assess their significance over time.

5.1.1.1 Emotion Overview

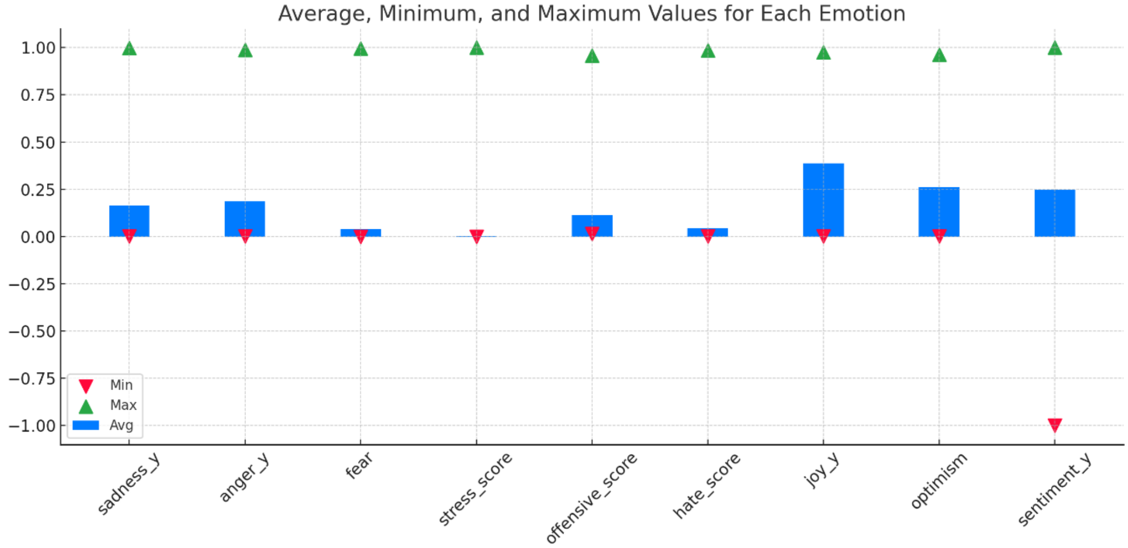
This section provides definitions and context for the emotions analyzed, each of which plays a significant role in understanding the emotional responses of Twitter users during the COVID-19 pandemic. The following descriptions elaborate on the specific emotional metrics:

- **Sadness**: This emotion is typically associated with tweets that express sorrow, grief, or a sense of loss. During the pandemic, sadness may be particularly reflected in tweets concerning personal loss, the broader impact of COVID-19 on communities, or reactions to distressing news events.
- **Anger**: Anger on Twitter is often seen in tweets expressing frustration, irritation, or criticism. In the context of the pandemic, anger could be directed at government responses or individual behaviors that are perceived as irresponsible.
- **Fear**: This emotion is expressed in tweets that convey worry, anxiety, or concern about the present or future situations. During a global health crisis, fear can be a common response to uncertainty about health risks, economic instability, or the social disruptions caused by the pandemic.
- **Stress Score**: This metric quantifies the level of stress evident in tweets. Higher stress scores may correlate with major pandemic events or changes in public health guidelines, reflecting the psychological strain experienced by users.
- **Offensive Score**: This score indicates the degree to which tweets contain language or themes that might be considered offensive or inappropriate. Understanding the offensive score helps determine the

presence of polarizing or sensitive content.

- **Hate Score**: This measures the intensity of hostility or animosity expressed in tweets. During the pandemic, increased hate scores might relate to xenophobia, racial tensions, or conflicts over pandemic handling.
- **Joy**: Represented in tweets that share positive experiences, celebrations, or gratitude. Even amidst the pandemic, moments of joy emerge in discussions about successful community actions, recovery stories, or personal achievements.
- **Optimism**: This emotion reflects a hopeful or positive outlook regarding future outcomes. Tweets with high levels of optimism during the pandemic might focus on vaccine developments, recovery rates, or supportive community actions.
- **Sentiment**: This general measure captures the overall emotional tone of a tweet, whether it is positive, negative, or neutral. Analyzing sentiment provides a broad measure of public mood and changes in attitude over time.

Emotional Range Analysis



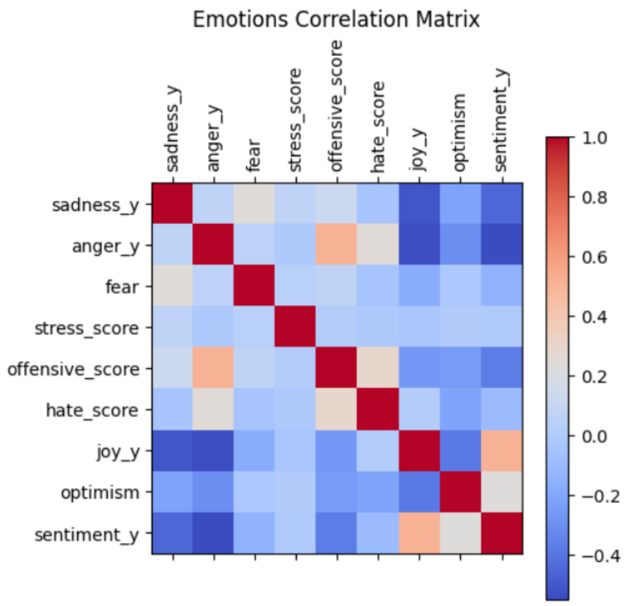
This chart offers a visual analysis of the average, minimum, and maximum values for various emotions and sentiment scores extracted from Twitter data. This visualization helps with understanding the depth of emotional responses across all emotions including sadness, anger, fear, stress, offensiveness, hate, joy, optimism, and overall sentiment during the COVID-19 pandemic.

The average values are indicated by blue bars, highlighting the central tendency of emotional expressions among Twitter users. These averages provide insights into the general mood as expressed in tweets. Notably, emotions such as joy and optimism, while exhibiting positivity on average, still show significant negative reaches, illustrating the public's complex emotions during the pandemic.

The minimum values, marked by red downward-pointing triangles, reveal the lowest points of emotional expression for each category. These points are important as they show the minimal presence or expression of each emotion. While the maximum values are shown by green upward-pointing triangles, showing the peak expression of each emotion. These peaks are particularly telling, as they reflect moments of intense emotional response.

The data reveals significant variability in the expression of emotions such as fear, sadness, anger, and stress, with a notably wide range between their minimum and maximum values, which points out the intense fluctuations in these emotions throughout the pandemic. Similarly, joy and optimism, while overall positive, also display considerable variability, reflecting the diverse and often volatile reactions among Twitter users as they respond to the evolving circumstances of the pandemic. On the other hand, offensive and hate scores, although generally lower on average, show peaks that are quite high, indicating sporadic but intense expressions of negativity. Furthermore, the overall sentiment across the dataset trends towards the positive, as suggested by the average score, yet the broad range from minimum to maximum values highlights a complex emotional landscape, with significant shifts in public sentiment that vary widely over the course of the pandemic.

Correlation Between Emotions



The correlation heatmap visualizes the relationships between a range of emotional and sentiment metrics from Twitter data, such as sadness, anger, fear, stress, offensive content, hate, joy, optimism, and overall sentiment. Each square in the matrix represents the correlation coefficient between pairs of these metrics, with the color intensity and hue indicating the strength and direction of the correlation.

Color Scheme: The heatmap uses a color gradient from red to blue. Shades of red indicate positive correlations, where the values of one metric tend to increase with the values of another. Shades of blue represent negative correlations, where one metric tends to increase as the other decreases. The intensity of the color reflects the strength of the correlation, with darker colors indicating stronger relationships.

Positive Correlations: Metrics such as offensive score and anger are strongly positively correlated, as shown by the orange squares. This suggests that tweets expressing higher levels of offensive content are likely to also express anger. Similarly, the correlation between overall sentiment and joy is also positive, indicating that more positive sentiment scores are associated with expressions of joy.

Negative Correlations: There are notable negative correlations, such as between joy and sadness, joy and anger, as well as overall sentiment with sadness and anger, highlighted by the dark blue squares. These indicate inverse relationships where higher expressions of joy or overall positive sentiment are associated with lower expressions of sadness and anger.

Neutral to Low Correlations: Some pairs of metrics, such as stress and fear or offensive and sadness, show very light shades, indicating weak or negligible relationships between these variables.

5.1.1.2 T-Tests

The purpose of conducting t-tests was to determine if there were statistically significant changes in the expression of various emotions on Twitter before and after the onset of the COVID-19 pandemic. The t-tests compared the mean scores of the emotions calculated from tweet data collected during two distinct periods, before and after the pandemic was declared. The results provide insights into how public emotional expressions shifted in response to the global crisis.

T-Test Results

emotion	before_mean	after_mean	t_statistic	p_value
fear	0.041	0.041	-20.258	0.000
anger_y	0.159	0.198	-586.796	0.000
joy_y	0.414	0.375	412.996	0.000
optimism	0.279	0.256	326.551	0.000
sadness_y	0.147	0.171	-415.465	0.000
sentiment_y	0.343	0.207	619.876	0.000
stress_score	0.001	0.001	-8.335	0.000
offensive_score	0.103	0.118	-486.884	0.000
hate_score	0.038	0.047	-696.687	0.000

The analysis revealed the following key changes in emotional expressions worldwide:

- **Fear**: The mean level of expressed fear remained unchanged (before: 0.041, after: 0.041), yet the t-statistic and p-value indicate a significant difference, likely due to a very small but statistically detectable change.
- **Anger**: There was a significant increase in the expression of anger, from a mean of 0.159 before to 0.198 after the pandemic started($t = -586.796$, $p < 0.001$), suggesting a heightened sense of frustration or irritation globally.
- **Joy**: Joyful expressions decreased significantly from a mean of 0.414 to 0.375 ($t = 412.996$, $p < 0.001$), reflecting a reduction in positive sentiments as the pandemic progressed.
- **Optimism**: Optimism also saw a decline, with mean scores dropping from 0.279 to 0.256 ($t = 326.551$, $p < 0.001$), indicating a more cautious or pessimistic outlook among Twitter users worldwide.
- **Sadness**: Expressions of sadness increased from a mean of 0.147 to 0.171 ($t = -415.465$, $p < 0.001$), highlighting an overall increase in sorrow and distress.
- **Sentiment**: Overall sentiment became more negative, decreasing significantly from 0.343 to 0.207 ($t = 619.876$, $p < 0.001$), which suggests a general downturn in mood across Twitter.
- **Stress**: Stress levels showed a minimal but statistically significant increase (before: 0.001, after: 0.001; $t = -8.335$, $p < 0.001$), pointing to increased anxiety or stress under pandemic conditions.
- **Offensive Content**: There was an increase in offensive content, from 0.103 to 0.118 ($t = -486.884$, $p < 0.001$), possibly due to heightened tensions and polarized sentiments.
- **Hate**: Hate expression also rose, with means increasing from 0.038 to 0.047 ($t = -696.687$, $p < 0.001$), suggesting an increase in hostile sentiments during the pandemic.

These t-test results indicate significant statistical change in all emotions following the outbreak of COVID-19, which is evident by the overall low p-values. The increases in anger, sadness, stress, offensive content, and hate, alongside decreases in joy, optimism, and overall positive sentiment, reflect a global emotional response to the challenges posed by the pandemic. These findings highlight the impact of global crises on public mood and behavior, offering important insights into the influence of the COVID-19 pandemic on the public.

5.1.1.3 Fixed Effect Regressions

The Fixed Effects Regression analysis was conducted to understand the impact of the COVID-19 pandemic on various emotions expressed by individual Twitter users over time. This approach controls for both observed and unobserved individual heterogeneity, isolating the effect of the pandemic from other individual-specific variations that do not change over time. By focusing on changes within individuals, the analysis provides insights into how users' emotional expressions shifted during the pandemic.

Regression Results

sentiment	coefficient	t_value	p_value	significance
fear	0.002	50.769	0.000	***
anger	0.006	72.637	0.000	***
joy	-0.022	-195.762	0.000	***
optimism	0.003	34.388	0.000	***
sadness	0.013	172.942	0.000	***
sentiment	-0.034	-125.216	0.000	***
stress score	0.000	26.609	0.000	***
offensive score	0.002	44.224	0.000	***
hate score	0.001	61.344	0.000	***

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

The analysis involved 156,092 unique Twitter users, yielding the following results for each emotion:

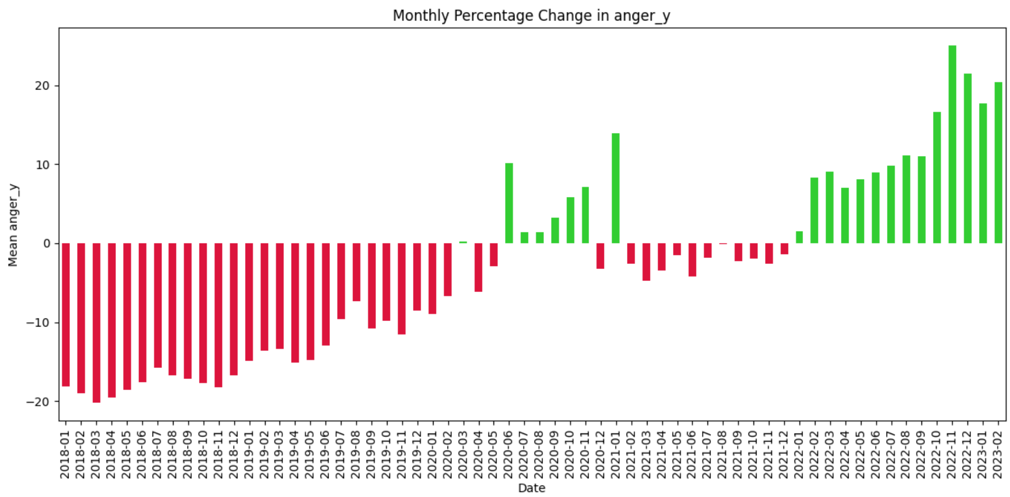
- **Fear**: There was a significant increase in the expression of fear, with a coefficient of 0.002 ($t = 50.769$, $p < 0.001$), indicating a rise in fear-related tweets as the pandemic progressed.
- **Anger**: Anger also showed a notable increase, with a coefficient of 0.006 ($t = 72.637$, $p < 0.001$), reflecting heightened feelings of anger among Twitter users during the pandemic.
- **Joy**: In contrast, joy experienced a significant decrease, as indicated by a coefficient of -0.022 ($t = -195.762$, $p < 0.001$). This suggests a decline in positive emotions expressed on Twitter as the pandemic unfolded.
- **Optimism**: Optimism saw a modest increase, with a coefficient of 0.003 ($t = 34.388$, $p < 0.001$), suggesting a slight rise in optimistic sentiments despite the overall challenges of the pandemic.
- **Sadness**: There was a substantial increase in sadness, with a coefficient of 0.013 ($t = 172.942$, $p < 0.001$), highlighting a significant growth in expressions of sadness among users.
- **Sentiment**: Overall sentiment on Twitter shifted negatively, with a coefficient of -0.034 ($t = -125.216$, $p < 0.001$), indicating a general downturn in mood among Twitter users.
- **Stress Score**: Stress levels showed a small but significant increase, evidenced by a coefficient of 0.000 ($t = 26.609$, $p < 0.001$).
- **Offensive Score**: Offensive content rose slightly, with a coefficient of 0.002 ($t = 44.224$, $p < 0.001$), indicating more frequent expressions of offensive sentiments.
- **Hate Score**: Similarly, hate expression increased, as shown by a coefficient of 0.001 ($t = 61.344$, $p < 0.001$), pointing to a rise in hate-related content on Twitter.

The results from the Fixed Effects Regression analysis show significant changes in the emotional expressions of Twitter users related to the COVID-19 pandemic. The overall increase in negative emotions (anger, sadness, stress, and hate) and the decrease in joy highlight the impact the pandemic has had on public sentiment. Conversely, the slight increases in optimism and offensive scores could reflect complex coping mechanisms and social dynamics in response to the pandemic.

5.1.1.4 Emotion Change Over Time

Exploring how public sentiment on Twitter has shifted throughout the pandemic reveals significant insights into the collective emotional response to a global crisis. By using bar plots to visualize the monthly percentage change of various emotions relative to their overall average, we can trace the changes in sentiments such as joy, sadness, anger, and optimism over time. These visualizations help us understand the broader impact of the pandemic on public mood and psychological well-being on a worldwide scale.

Anger

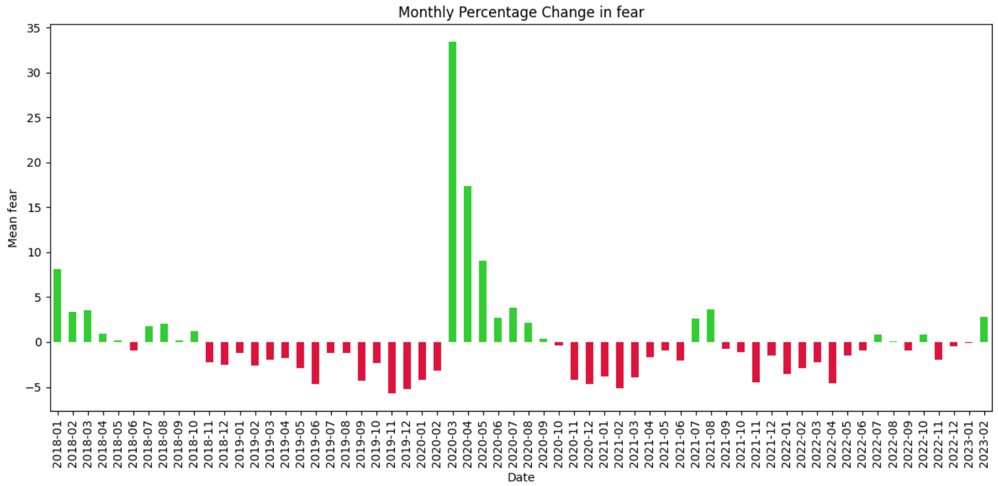


The analysis of anger trends, as depicted in the monthly percentage change bar plot, reveals an increase in expressions of anger coinciding with major global and political events during the COVID-19 pandemic. Initially, anger shows an escalating pattern leading up to the onset of the COVID-19 pandemic [14], which sharply spiked in June 2020. This spike corresponds with two significant events: the global death toll from COVID-19 surpassing 500,000 [33] and the international protests sparked by the tragic death of George Floyd [34]. These events collectively contributed to a surge in anger, highlighting a widespread response to both the pandemic's toll and social injustice.

In January 2021, another significant increase in anger was observed, which can be primarily attributed to the US presidential elections and the subsequent Capitol assault [35]. Given the substantial representation of tweets originating from the USA in our dataset, these events have a considerable influence on the overall emotional observed.

Following these events, a sustained increase in expressions of anger is observed from February 2022 onwards, which aligns with the onset of Russia's invasion of Ukraine [36]. Furthermore, this increase in anger sees new spikes around October 2022 with the acquisition of Twitter by Elon Musk [37].

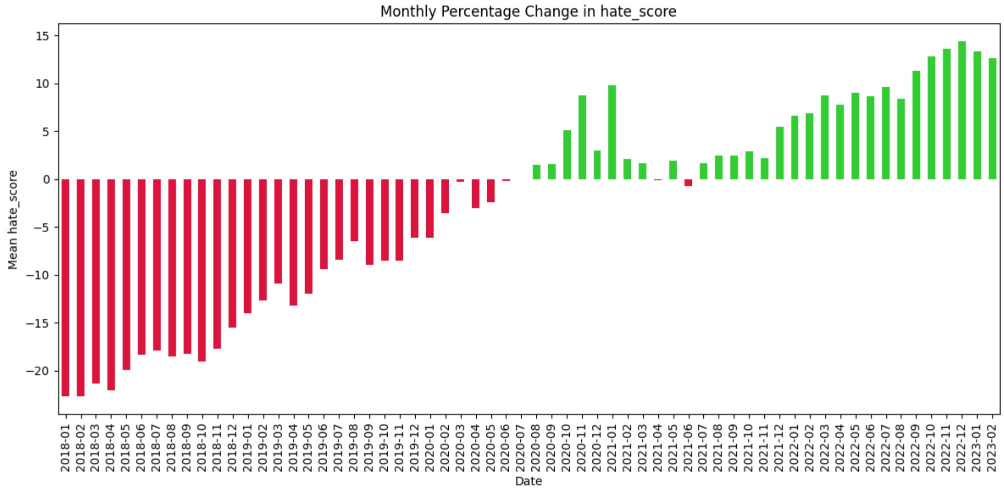
Fear



The bar plot illustrating the monthly percentage change in expressions of fear on Twitter initially shows a surge in fear, coinciding with the global outbreak's onset. The persistence of elevated fear during the early months of the pandemic emphasizes the continued anxiety about health risks and the effectiveness of the initial response measures.

A notable spike in fear is observed in July 2021, corresponding with the emergence and rapid spread of the Delta variant of the coronavirus [38]. This variant, known for its increased transmissibility and potential to evade immunity, reignited global concerns, particularly regarding the efficiency of existing vaccines against variant strains.

Hate Score

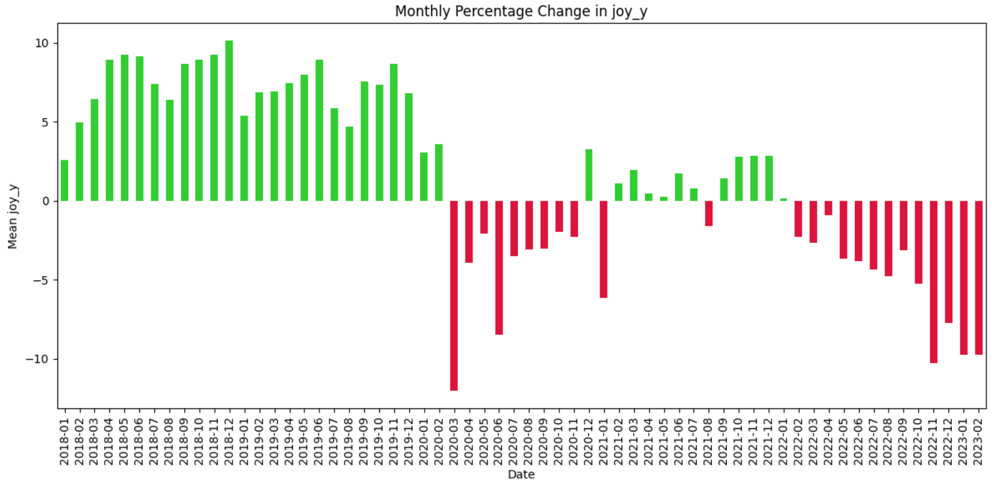


The trajectory of hate score on Twitter reflects a gradual increase in public sentiment over time. Initially, expressions of hate were below average, suggesting a period of relative calm or restraint in online interactions. However, as the pandemic unfolded, a significant change occurred with hate scores gradually rising above average.

This upward trend in hate expression first became notable around August 2020 and intensified by October 2020. These months coincided with a global resurgence in COVID-19 cases, which led to renewed lockdowns and public health restrictions [39]. The associated social and economic strains likely contributed to the observed increase in hate, as frustrations and societal divisions came to the forefront.

The pattern of steadily increasing hate expressions persisted, eventually accelerating with the outbreak of the Ukraine conflict in early 2022 and the Twitter acquisition later in the year. These events further amplified global tensions, reflected in the continuous rise in hateful sentiments on social media. The increasing hate scores during these periods highlight how external crises can significantly influence public emotions, driving more frequent expressions of hate in online platforms.

Joy



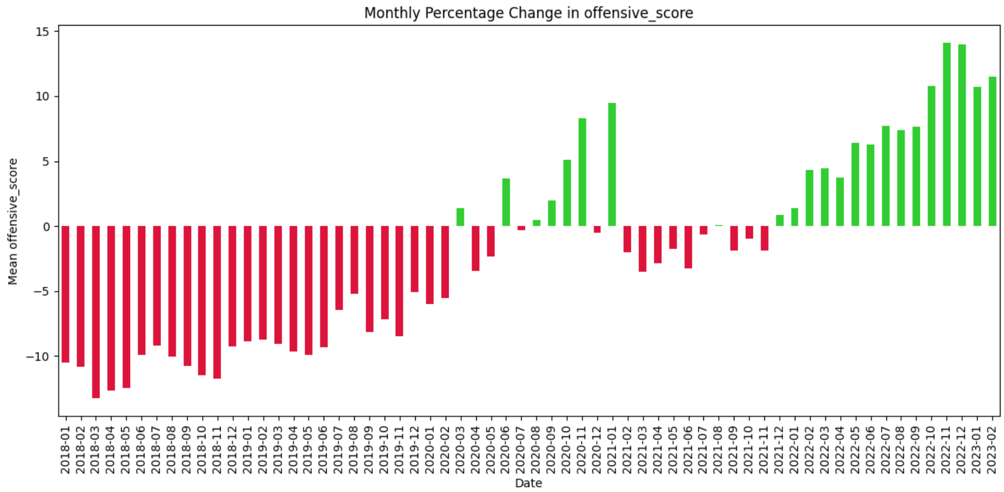
Expressions of joy on Twitter showcase a dynamic response to global events during the COVID-19 pandemic. Initially, joy exhibited a positive trend, reflecting a period of relative normalcy before the full worldwide impact of the pandemic. As the pandemic started, a negative trend in joy began, showing global concern and uncertainty.

This negative trend persisted through most of 2020, reaching its lowest the month the outbreak happened. However, December 2020 marked a turning point with the commencement of COVID-19 vaccination campaigns across various countries [40]. The authorization and distribution of vaccines like Pfizer and Moderna brought a sense of hope, reflected in a significant increase in expressions of joy. This resurgence indicates a collective relief as effective vaccines promised a potential end to the pandemic.

Despite this optimistic rebound, joy's upward trajectory faced a disruption in January 2021, coinciding with major political events in the United States, including the presidential elections and the Capitol assault.

Following these disturbances, joy began to recover once more, climbing steadily as the initial shock of political events waned and vaccination efforts continued to expand. However, this recovery was again interrupted by two major events, the invasion of Ukraine in early 2022, which introduced new uncertainties and challenges, and the acquisition of Twitter, impacting global sentiment and once again influencing the expressions of joy observed on the platform.

Offensive Score

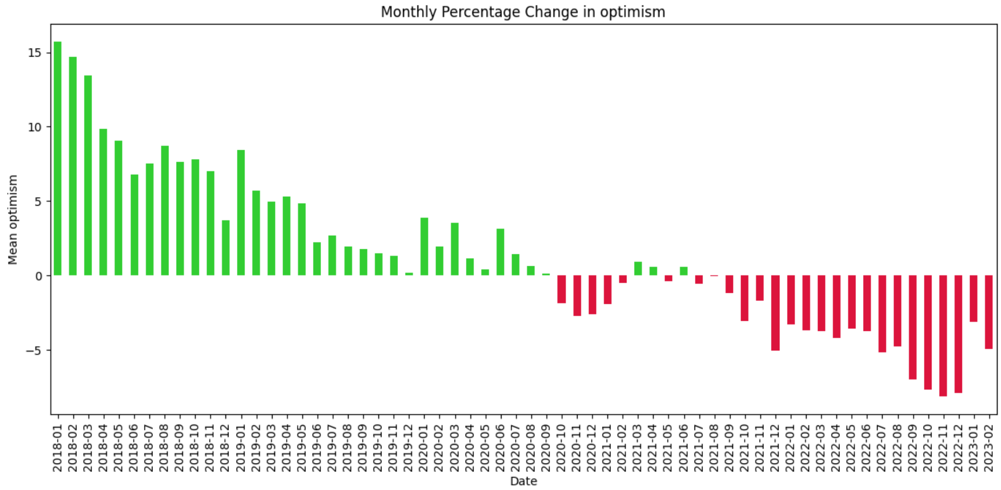


The monthly change in offensive scores on Twitter showcases a distinct pattern influenced by global events during the pandemic. Initially experiencing a steady decline, the offensive content on Twitter began to surge notably with the onset of COVID-19. This increase in offensiveness first peaked as the global death toll reached 500,000 and during widespread Black Lives Matter protests, reflecting the societal tensions and heightened emotional discourse online.

As the pandemic continued, further spikes in offensiveness are evident during periods of renewed public health measures and lockdowns between August 2020 and October 2020, which contributed to growing public frustration and discontent expressed through social media. The trend of increased offensiveness persisted into January 2021, intensified by the storming of the Capitol Building in the USA.

In early 2022, the offensive score spiked again, coinciding with the invasion of Ukraine. This period marked another rise in global anxiety and debate, impacting the tone of discussions on Twitter. After that, further spikes can be observed with the change in Twitter's ownership around October 2022.

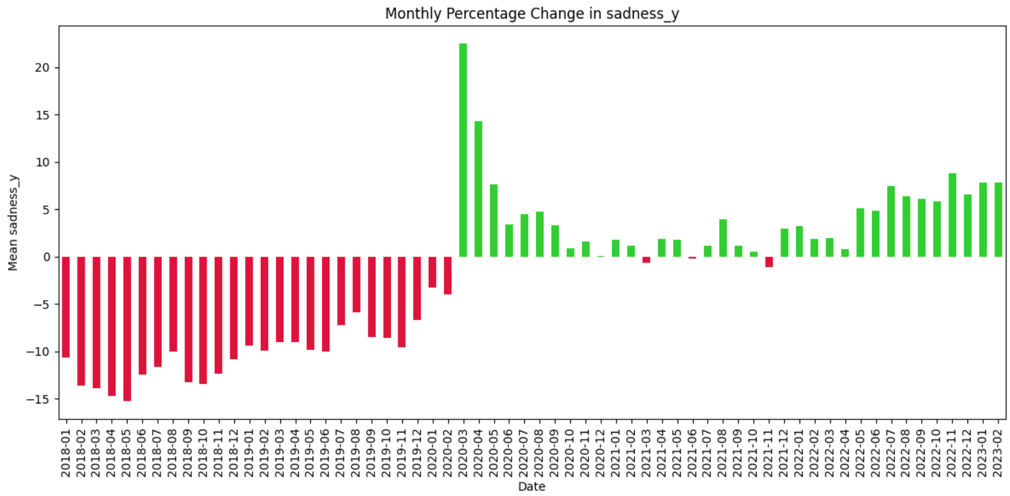
Optimism



Initially, optimism showed resilience, maintaining a relatively stable or less impacted state as the pandemic began. This initial phase suggests a collective hope or positive outlook possibly driven by a belief in quick resolutions or effective management of the situation.

However, as the pandemic prolonged, marked by various challenges including new waves of infections and the emergence of new variants, a decline in optimism is observable. This trend reflects a growing realization of the pandemic's severity and complexity. The continued negative trajectory of optimism becomes more pronounced with the geopolitical tensions marked by the invasion of Ukraine and the acquisition of Twitter.

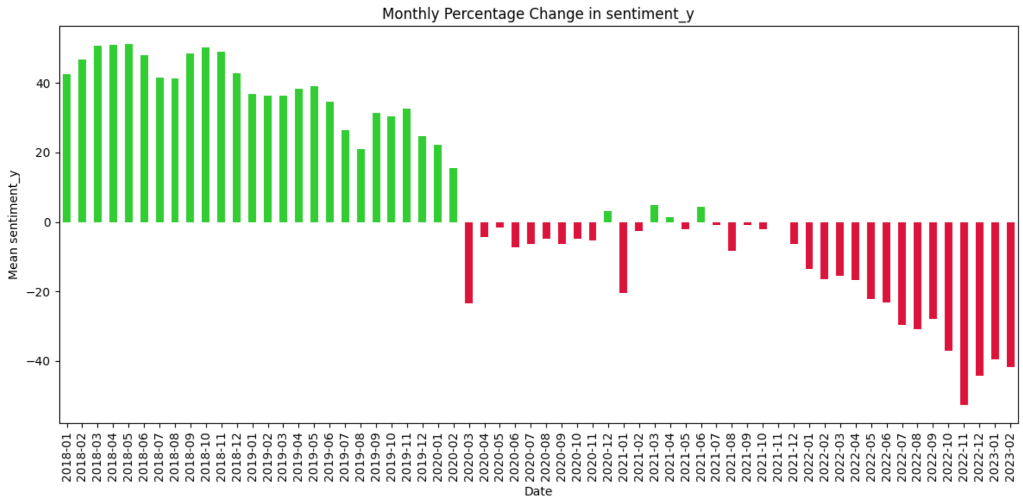
Sadness



An immediate and sharp increase in sadness is evident as the pandemic begins. This initial surge in sadness highlights the emotional response to the sudden and widespread impact of the virus, affecting everyone's lives, and global economies.

After this initial spike, there appears to be a period of relative stability where sadness levels off. This stabilization might suggest a phase of adaptation where individuals and communities begin to adjust to the new normal. However, this is disrupted by the invasion, triggering another significant and persistent rise in expressions of sadness. The resurgence of heightened sadness aligns with the uncertainty and additional strain on global stability caused by the invasion, emphasizing the worldwide distress caused by successive global crises.

Sentiment

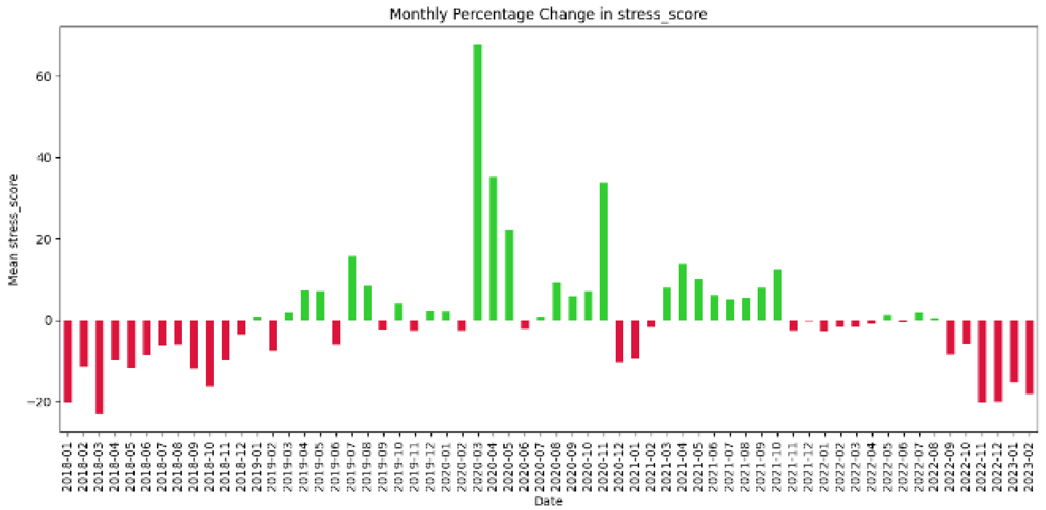


Prior to the pandemic, the sentiment remains consistently above average, indicating a generally positive or stable public mood. However, as the pandemic starts, there's a shift, with noticeable spikes in sentiment reflecting the collective emotional reaction to the initial shock and the unfolding global crisis.

As the pandemic progresses, specific events cause distinct responses in sentiment. The beginning of the pandemic is characterized by a dramatic drop, signaling widespread anxiety and concern as countries are faced with the implications of the virus. Another significant dip occurs during the storming of the Capitol Building in the United States, a period marked by political turmoil and uncertainty, which further influences public sentiment.

The most pronounced and sustained decrease in sentiment occurs following the invasion of Ukraine. This period not only extends the trend of negative sentiment but deepens it, with even bigger negative spikes being observed around the time Twitter was sold.

Stress Score



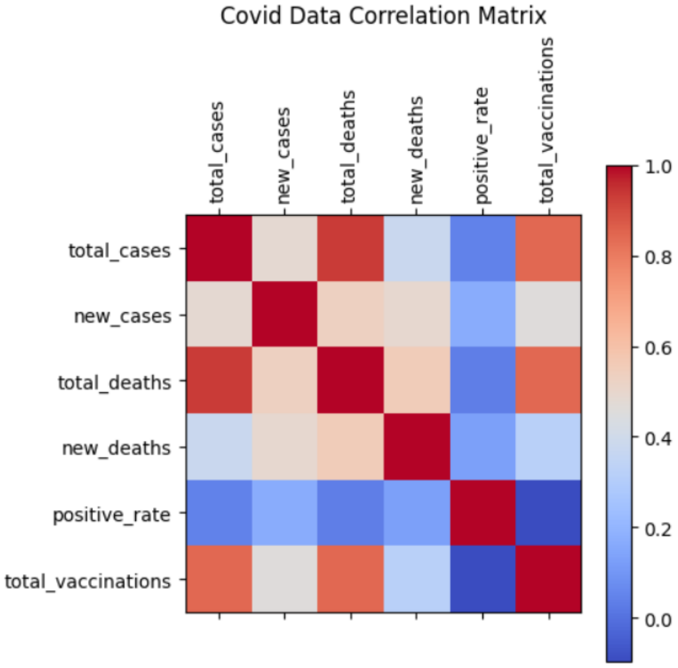
Analyzing the stress score graph reveals a subtle yet discernible pattern in the fluctuations of stress levels as captured on Twitter over time. The graph shows periods of relatively low variability in stress scores, indicating minor differences from month to month. This low variability suggests that while there are peaks, the changes in stress levels are not as pronounced or as volatile as those seen in other emotional metrics like sadness or anger.

Noteworthy, however, are the distinct peaks that do occur, particularly around specific events that align with significant global or national crises. These peaks suggest that while the overall variability in stress scores may be low, certain events trigger noticeable spikes in stress. For example, the tallest peak observed correlates with the initial outbreak of COVID-19, a time of high uncertainty and anxiety which would naturally lead to increased stress levels.

Subsequent smaller peaks may correspond to ongoing pandemic developments or other stressful news events, but without the same magnitude of impact as the initial outbreak. The relatively stable stress levels post these events could indicate a normalization of higher stress, or it might reflect limitations in the data's sensitivity to capture more gradual increases in stress levels over time.

5.1.2 Covid Data Correlations

Exploring the relationships among various COVID-19 epidemiological data points provides key insights into how different aspects of the pandemic are interlinked. The heatmap displayed offers a visual representation of these correlations, showcasing how different pandemic metrics might influence one another across various contexts.



The heatmap uses a color gradient from red to blue to indicate the strength and direction of correlations between pandemic metrics such as total cases, new cases, total deaths, new deaths, positive rate, and total vaccinations. Red hues suggest positive correlations where metrics tend to increase together, whereas blue hues indicate negative correlations, suggesting that the rise in one metric might correspond with a decrease in another.

Positive Correlations

There is a strong positive correlation between total cases and total deaths, indicating that regions with higher infection rates also experience higher mortality rates. Similarly, new cases and new deaths are positively correlated, showing that spikes in case numbers often coincide with increases in deaths.

Total vaccinations are positively correlated with both total cases and total deaths. This correlation may initially seem counterintuitive but can reflect higher absolute numbers in heavily populated or heavily affected regions that are prioritizing extensive vaccination campaigns.

Negative Correlations

There is a noticeable negative correlation between the positive rate and both total cases and total deaths. This suggests that as the total number of cases and deaths increases, the proportion of positive tests tends to decrease, possibly due to wider testing that includes more individuals with milder or no symptoms.

A significant negative correlation is observed between total vaccinations and the positive rate, indicating that higher vaccination rates contribute to a lower proportion of positive tests, reflecting the effectiveness of vaccines in controlling the spread of the virus.

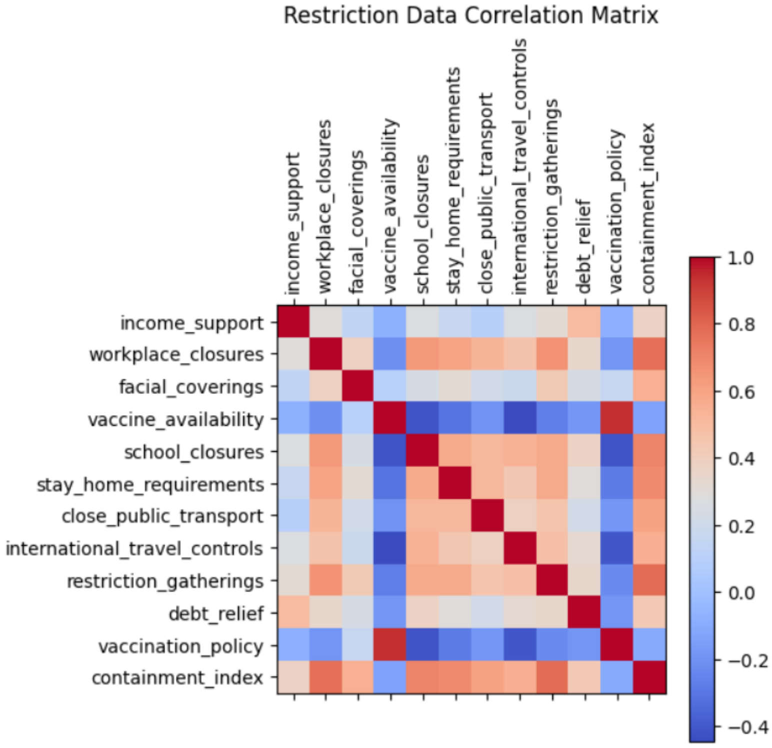
Minor Correlations

The relationship between new deaths and new cases shows almost no correlation, suggesting that their relationship can vary significantly depending on local healthcare responses and the severity of cases.

The correlation between total vaccinations and new cases is also nonexistent, indicating that while vaccination efforts are ramping up, new cases are not decreasing as uniformly, likely influenced by factors such as vaccine distribution challenges or emerging virus variants.

5.1.3 Covid Restrictions Correlations

Exploring the relationships among various COVID-19 restriction measures provides key insights into how different policies may complement or counteract each other. The heatmap displayed offers a visual representation of these correlations, highlighting how the implementation of one measure might be associated with or differ from others across different regions.



The heatmap has a color gradient ranging from red to blue to indicate the strength and direction of correlations between restriction measures such as income support, workplace and school closures, mask mandates, vaccination policies, and more. Red hues suggest positive correlations where measures tend to be implemented in conjunction, whereas blue hues indicate negative correlations, suggesting that the presence of one measure might reduce the likelihood of another being enforced simultaneously.

Positive Correlations

There is a strong positive correlation between restrictions on gatherings and the closure of workplaces, indicating that when one is enforced, the other is likely also to be in effect, reflecting a comprehensive approach to limit public interactions.

The containment index, which measures the overall strictness of government actions intended to contain the spread of the virus, shows a strong positive correlation with several major social distancing measures. This includes workplace and school closures, restrictions on gatherings, and stay-home requirements. These correlations indicate that higher values of the containment index are associated with more strict enforcement of social distancing policies. The containment index serves as an indicator of how aggressively a region is responding to the pandemic through physical distancing efforts. As such, when the index is high, it typically means that extensive social distancing measures are in place, reflecting a proactive approach to minimizing public interactions and reducing virus transmission rates.

Negative Correlations

As vaccine distribution progresses, there is a significant negative correlation between vaccine availability (and associated vaccination policies) and the same social distancing measures linked to the containment index. This pattern suggests that as vaccination rates increase, the reliance on strict social distancing measures decreases, likely due to the growing immunity within the community, which allows for a gradual relaxation of restrictions.

Minor Correlations

Measures like income support and facial coverings do not show strong correlations with other restrictive measures. These trends suggest that the implementation of financial aid and mask mandates is likely driven by specific situational needs rather than a coordinated response with other restrictions. This implies that while certain policies are universally applied, others like income support and facial coverings are adapted more flexibly and independently, based on local economic conditions and public health priorities.

5.1.4 Topic Analysis

The topic analysis delves into the thematic content of tweets throughout the COVID-19 pandemic, identifying and analyzing the major topics that dominated discussions on Twitter. By exploring these topics, we can understand not only what subjects gathered the most attention but also how they contributed to the change of public sentiment. This analysis will reveal the most popular topics overall as well as those that peaked in interest at different times throughout the pandemic.

5.1.4.1 Topics Overview

Examining the major topics discussed on Twitter, this subsection outlines the themes that dominated conversations during the pandemic. It identifies which topics were most prevalent and how these preferences shifted over time, reflecting changes in public interest and concerns as the pandemic evolved.

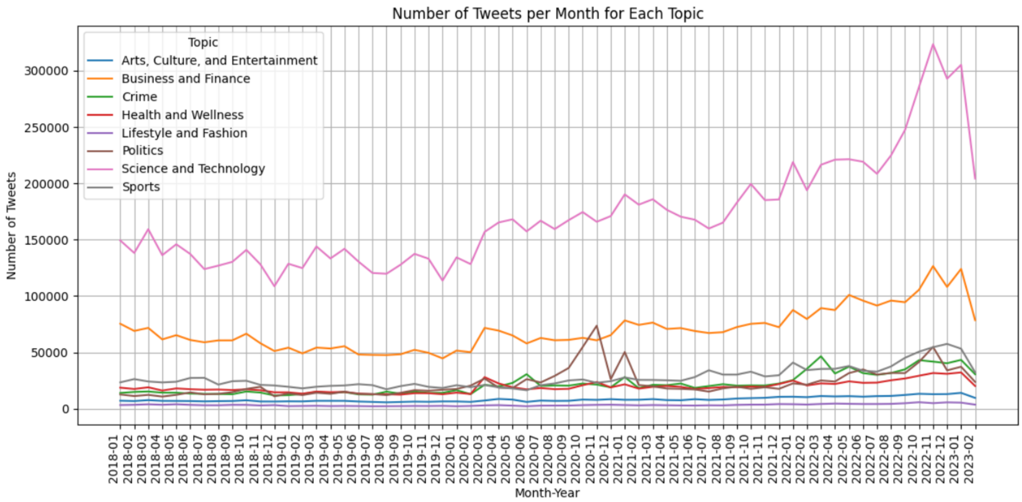
Most Popular Topics

Topic	Number of tweets	Percentage
Science and Technology	10629000	49.8%
Business and Finance	4341772	20.3%
Sports	1706526	8.0%
Politics	1399539	6.6%
Crime	1339083	6.3%
Health and Wellness	1191338	5.6%
Arts, Culture, and Entertainment	513276	2.4%
Lifestyle and Fashion	206517	1.0%

The analysis of topics discussed during the COVID-19 pandemic highlights a concentrated focus on certain key areas. Science and Technology emerged as the predominant theme, making up for almost half of the tweets with 10,629,000 tweets, representing 49.8% of all topic-centered tweets. This was closely followed by Business and Finance, accounting for 4,341,772 tweets or 20.3% of the total.

Out of the total 21,327,051 tweets, a substantial 60% (12,901,785 tweets) originated from the USA, highlighting the significant influence of American discourse on these topics. The overwhelming focus on Science and Technology, followed by Business and Finance, accounted for 70% of all indexed tweets, emphasizing the critical concern for these areas during the pandemic period. This distribution not only reflects the global urgency for scientific solutions and economic implications during the crisis but also shows the varied interests and concerns of the Twitter community during a time of crisis.

Monthly Tweets per Topic



The chart illustrating the number of tweets per month for each topic provides a detailed view of how discussion volumes have evolved over the course of the pandemic, particularly highlighting variations in public interest across different topics.

Science and Technology remains a standout topic, demonstrating a steady increase in tweet volume since the pandemic's onset. This uptrend sees a significant increase around mid-2022, reflecting heightened public engagement possibly due to emerging technological innovations or breakthroughs in medical science related to the pandemic. The consistent interest in this topic accentuates the central role of technology and science in addressing the challenges posed by COVID-19.

Across all topics, there is a noticeable general increase in tweet volumes after mid-2022. While this could partly be attributed to an overall increase in

Twitter activity, it also suggests a growing engagement with various subjects as the pandemic continued to influence global discourse.

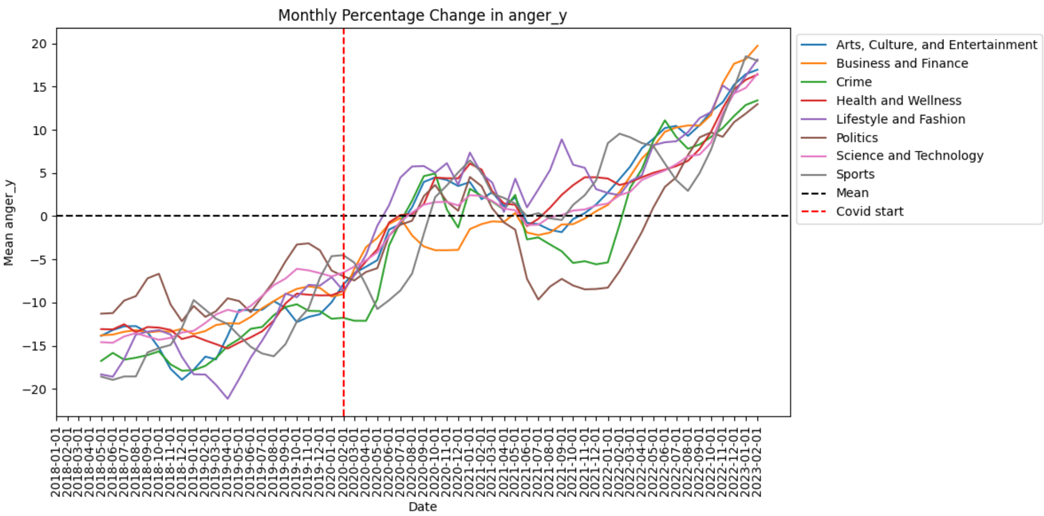
Politics shows distinct spikes in activity, particularly around November 2020, aligning with the U.S. presidential election, and extending into January 2021, during significant political events including the Capitol riots. Given that 60% of the dataset's tweets are from the USA, these spikes show the impact of national events on global social media conversations.

This analysis confirms that while some topics maintained steady interest over time, others were subject to fluctuations because of specific events, showing how dynamic public interest and engagement are on social media platforms during significant global and national events.

5.1.4.2 Change in Emotions Over Time by Topic

Analyzing the emotional responses associated with specific topics discussed on Twitter reveals how public sentiment shifted in response to the pandemic's developments. This analysis connects the dots between major events or news and their emotional impact, highlighting how specific topics of discussion influenced collective mood and attitudes over time.

Anger



Across all topics, there is a gradual increase in anger, suggesting a mounting frustration or contentiousness in discussions as the pandemic persisted. This trend reflects the growing public dissatisfaction and stress related to ongoing global challenges.

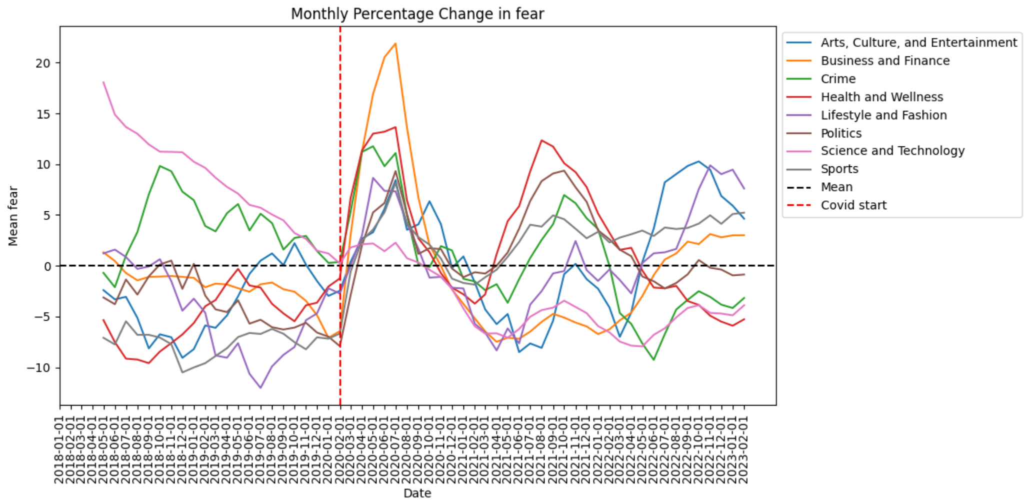
A temporary decline in anger is observable in the summer and fall of 2021 across several topics. This dip could potentially align with periods of pandemic fatigue where discussions might have momentarily shifted towards more hopeful topics or when significant progress in vaccine distribution was achieved, momentarily easing public anxieties.

The topic of Politics stands out with the most significant fluctuations in anger, particularly evident during key political events such as elections, legislative actions, and other government-related controversies. The sharp peaks in anger within political discussions underscore the high tensions and polarized opinions prevalent during these times.

Business and Finance also shows increases in anger, likely reflecting public sentiment regarding economic policies, unemployment, financial crises, and their management during the pandemic.

Interestingly, even discussions under Health and Wellness, a topic crucial during a health crisis, show an increase in anger, possibly due to debates over health measures, vaccine mandates, and the handling of public health directives.

Fear



There is a noticeable spike in fear across most topics during the early months following the pandemic's outbreak, indicated by the sharp rises near the COVID-19 start line. This widespread increase reflects the global uncertainty and anxiety triggered by the sudden emergence and rapid spread of the virus.

Particularly for Business and Finance and Health and Wellness, fear levels show significant peaks. In Business and Finance, this could be attributed to economic uncertainties, market disruptions, and concerns about job security and financial stability. Health and Wellness discussions reflect fears about the virus itself, the effectiveness of health measures, and the implications for personal and public health.

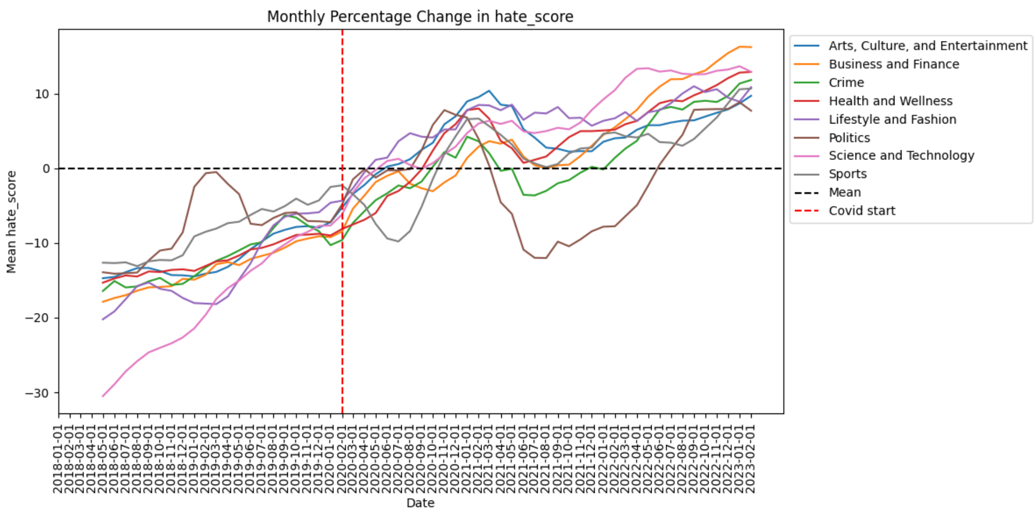
Topics like Arts, Sports, and Lifestyle exhibit a resurgence in fear towards late 2022. This could be linked to the ongoing challenges in these sectors, possibly due to new COVID variants affecting public gatherings, events, and general lifestyle adjustments.

The topic of Science and Technology shows a distinct pattern, largely remaining unaffected by the initial surge of fear. This resilience might come from the sector's role in providing solutions, such as vaccine development and technological innovations for combating the pandemic, which may have contributed to a more stable or even optimistic view.

Business and Finance is the most impacted topic by fear, discussions in this area are likely centered on the immediate and long-term economic repercussions of the pandemic, reflecting concerns over financial crises and their broader societal effects.

Health and Wellness's significant fear response emphasizes the direct impact of the health crisis on individual and community health perspectives, driven by the information about the virus's spread and containment.

Hate Score

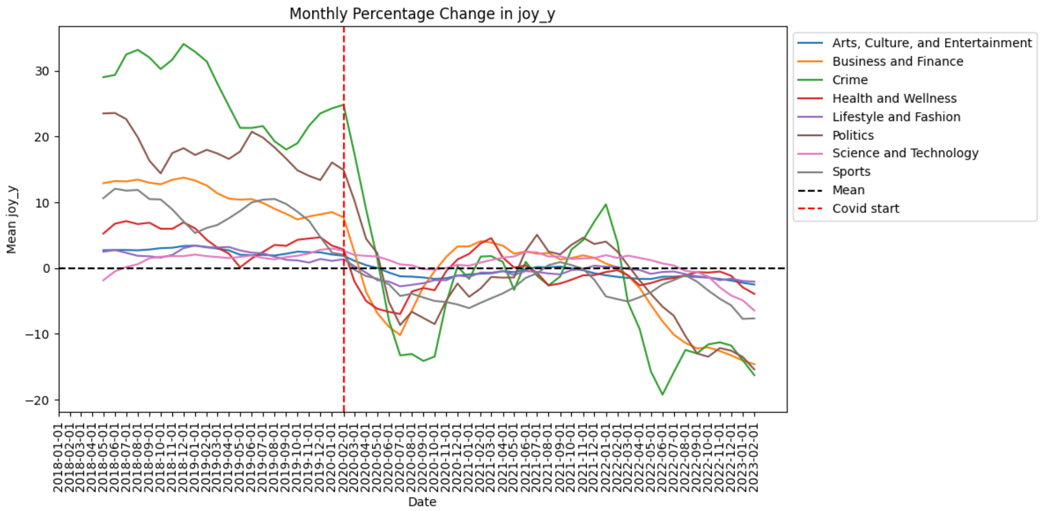


The trends in hate score across various topics closely mirror those observed in anger, exhibiting similar patterns of gradual increases and notable fluctuations tied to specific events. As with anger, all topics show an overall rise in hate, followed by a temporary decline during the summer and fall of 2021. This parallel suggests a consistent alignment between expressions of anger and hate, where topics that provoke anger also tend to provoke hateful comments.

Similar to the trends in anger, the topic of Politics stands out in the hate score analysis, displaying the largest and most frequent shifts. This correlation underscores how political events and controversies can cause both anger and hate on social media.

The general increase in hate across all topics further aligns with the anger analysis, indicating that the pandemic and its associated challenges have not only heightened frustrations but also intensified expressions of hate.

Joy



The overall trend shows a sharp decline in expressions of joy coinciding with the pandemic's outbreak, marked by the red dashed line indicating the start of COVID-19. This decline reflects the widespread impact of the pandemic on global mood and outlook.

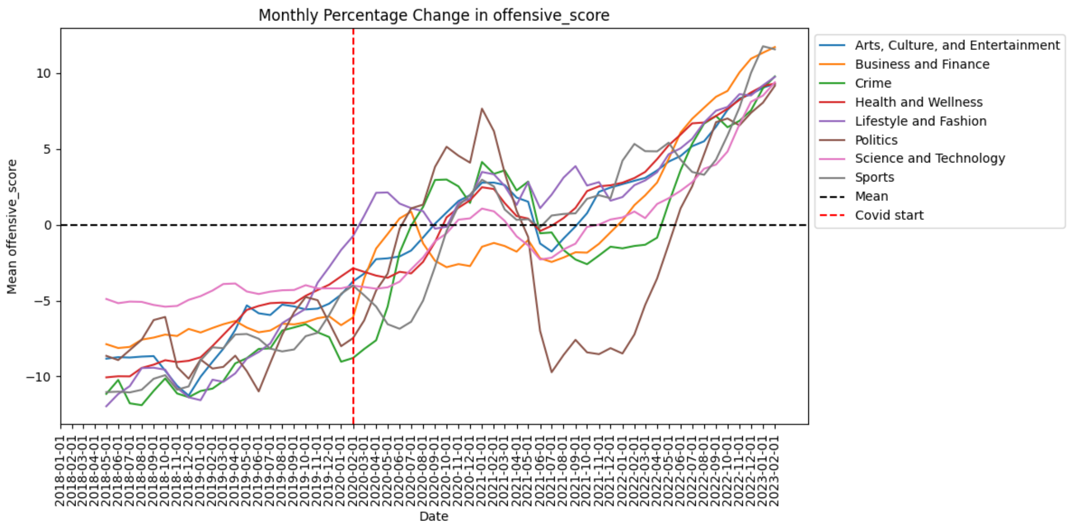
Tweets related to Science and Technology show resilience in maintaining a relatively stable joy level, even experiencing increases at certain points. This stability might be attributed to ongoing technological advancements and scientific breakthroughs during the pandemic, which could have provided hope and optimism.

Arts, Culture, and Entertainment, and Lifestyle and Fashion exhibit only a slight decrease in joy, suggesting that while affected, discussions in these areas retained a degree of positivity. This could be due to the role of arts and lifestyle content in providing a diversion from the pandemic's stresses.

Crime, Politics, and Business and Finance are the topics that show the most substantial decreases in joy. The nature of these discussions often involves more serious or contentious issues, which may have been particularly heightened during the pandemic due to increased economic uncertainties, political unrest, and social issues, thereby heavily impacting the associated emotional tone.

Over time, the recovery trajectories vary by topic, indicating that the nature of the discussion significantly influences how quickly or slowly joy levels rebound. For instance, more utilitarian or essential topics like Business and Finance struggle to regain positive sentiment, likely reflecting ongoing economic challenges.

Offensive Score



When it comes to the offensive score, there is a clear upward trend across all topics, reflecting a general rise in contentious or provocative language on Twitter as the pandemic unfolded. This trend may be attributed to heightened emotions and stress within the public during this period.

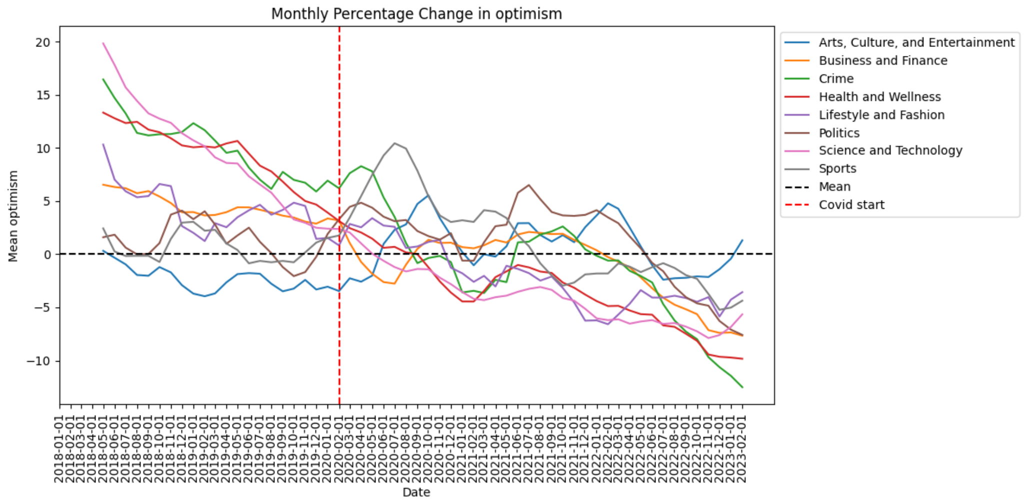
A notable peak in offensiveness occurs across most topics towards the end of 2020. This timing correlates with several major global events, including the announcement and initial rollouts of COVID-19 vaccines, which may have sparked debates and controversy, thereby increasing offensiveness in related discussions.

Business and Finance peaks in offensiveness during the summer of 2020, possibly reflecting intense reactions to economic policies, market uncertainties, or corporate responses to the pandemic, which were debated during the early phases of global lockdowns [41].

As anticipated, offensiveness in political discussions shows significant turbulence, with sharp peaks during the U.S. presidential election in late 2020 and subsequent events such as the storming of the Capitol Building. These political events likely fueled increases in offensive content. A resurgence in offensive scores is also observed with the onset of the Ukraine invasion in early 2022, indicating that geopolitical conflicts continue to incite offensive language in political discourse.

Furthermore, there is a noticeable decline in offensiveness following the U.S. election and the Capitol storming, suggesting a temporary resolution or fatigue in public debates following these key political milestones.

Optimism



Certain Topics like Science and Technology, Health and Wellness, and Crime show a consistent downward trend in optimism. This could be linked to ongoing challenges within these sectors, for example, Science and Technology facing the pressure of rapid innovation demands, Health and

Wellness burdened by the health crisis implications, and Crime possibly reflecting societal unrest.

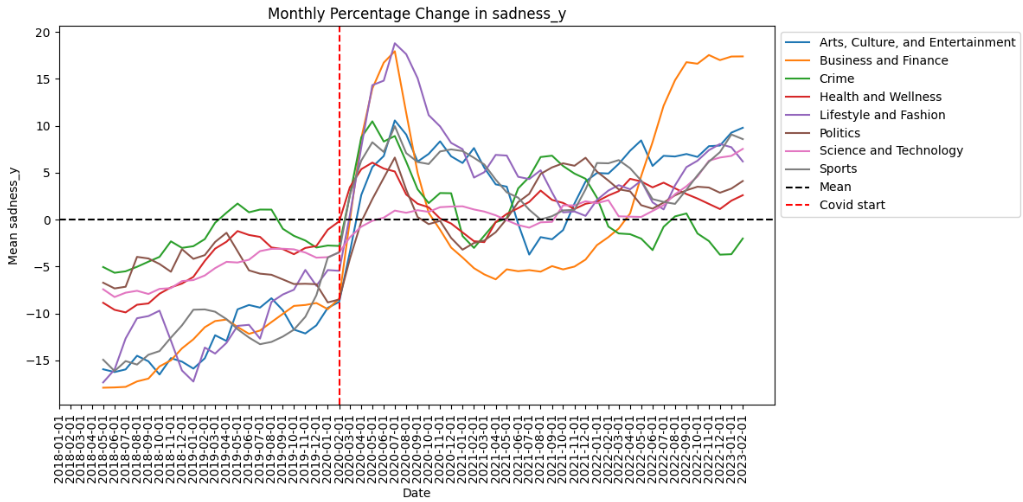
Optimism within the Arts, Culture, and Entertainment sector remains relatively stable over the long term. This resilience might come from the sector's role in providing escapism and relief during stressful times, helping maintain a level of positive sentiment.

Business and Finance shows a notable reaction at the pandemic's onset, possibly reflecting immediate concerns about economic disruption which then stabilizes as businesses adapt to new realities.

The optimism in Sports initially increases despite significant events like the cancellation of the 2020 Olympics [42]. This initial rise could reflect a hopeful sentiment about the potential for rapid resolution to the pandemic or adaptiveness of the sports sector through virtual and modified events. A later increase in optimism in the summer of 2021 coincides with the rescheduled Tokyo Olympics, highlighting specific events' capacity to boost public morale and optimism.

Most topics do not show drastic changes in optimism at the pandemic's onset, suggesting either a delayed emotional response or a buffer effect where the full impact took time to manifest in public sentiment.

Sadness



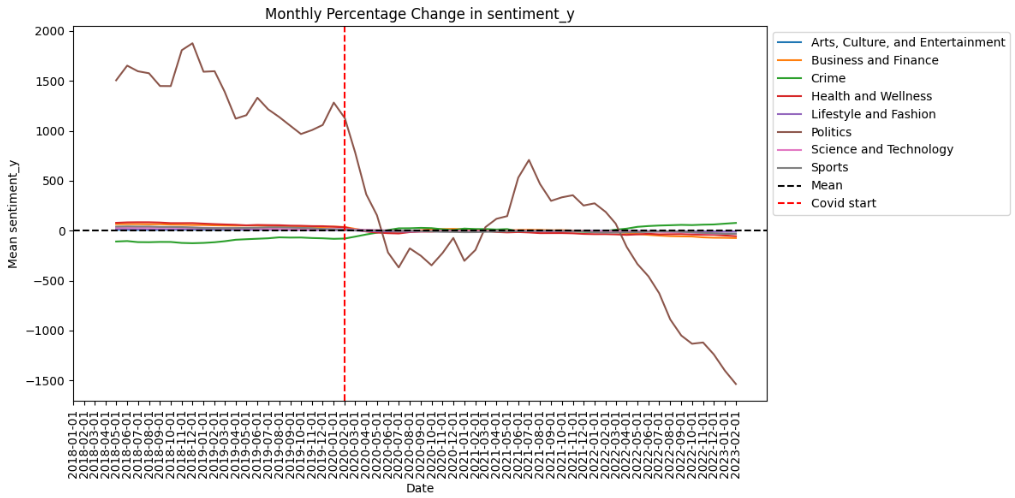
Consistent with the start of the pandemic, there is an increase in sadness across nearly all topics. This surge aligns with the global shock and the widespread disruption caused by the initial spread of COVID-19, reflecting a collective mourning over lost lives, disrupted lives, and uncertainty about the future.

While some topics like Science and Technology show a gradual increase in sadness, which likely correlates with ongoing challenges and pressures in these fields, other topics exhibit more pronounced fluctuations.

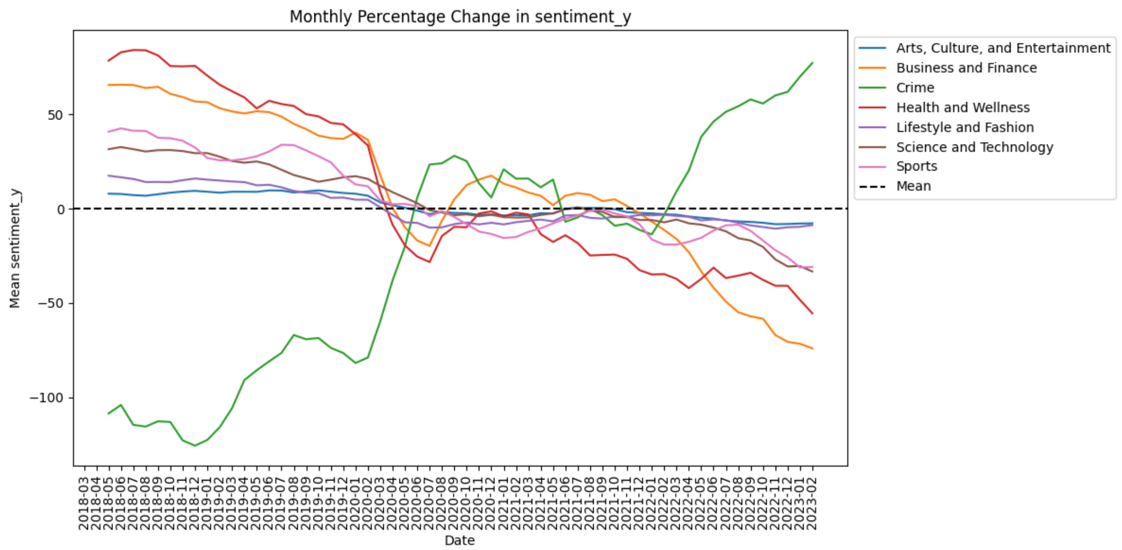
Business and Finance and Lifestyle and Wellness show significant peaks in sadness during the summer of 2020. This spike could be attributed to the intense strain on businesses and individuals' well-being during the first wave of the pandemic and initial lockdowns. The drop in sadness post-summer 2020 suggests a brief period of adaptation or relief as lockdowns eased and businesses started to reopen.

A notable resurgence in sadness in early to mid-2022 for these topics suggests the influence of other factors, possibly related to economic downturns, renewed health concerns, or other societal crises that may have reignited concerns and exacerbated feelings of sadness.

Sentiment



The sentiment associated with political discussions shows substantial volatility. An initial sharp decline is observed at the start of the COVID-19 pandemic, reflecting the intense and polarized nature of political discourse during this period. Although there is a brief improvement in mid-2021, the trend over time remains predominantly negative, hitting an all-time low by February 2023.

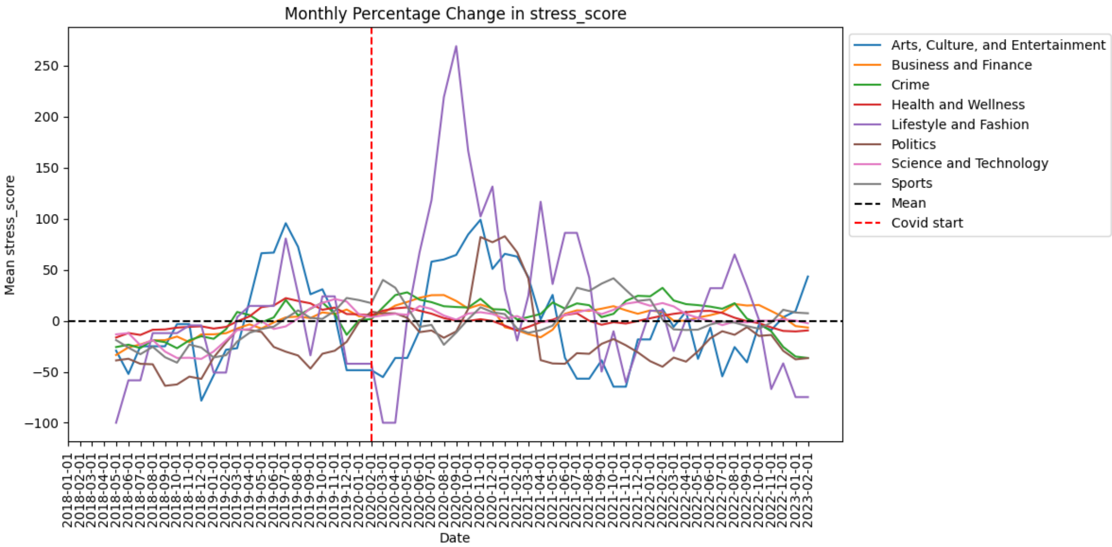


Due to the turbulent changes in sentiment for Political tweets, a graph without them was created so the rest of the topics can be examined.

With political tweets excluded, the sentiment trends across other topics appear more uniform but still generally declining. Notably, topics like Business and Finance, and Health and Wellness experience a decline in sentiment at the pandemic's onset, likely due to immediate impacts on economic conditions and public health fears.

An interesting divergence is seen in discussions related to Crime, where sentiment increases over time. This counterintuitive trend might be influenced by the public's reaction to negative events, possibly reflecting a rallying spirit or a community's unified stance against crime during crises such as the pandemic and the Ukraine war.

Stress Score



The stress scores across all topics display considerable turbulence, which may be attributed to the binary nature of the stress index used in this analysis (0 or 1). This scoring method captures whether stress is present or not, leading to sharp fluctuations in the calculated monthly changes.

Topics such as Lifestyle and Wellness, Arts, Culture and Entertainment, and Politics show larger variations in stress levels month-to-month. These areas likely reflect immediate reactions to ongoing events, with public discourse in these fields being directly impacted by changes in social conditions, policy decisions, and cultural events.

On the other hand, topics like Business and Finance, Science and

Technology, and Health and Wellness exhibit more stability in stress levels. The steadier trend in these areas could suggest that discussions within these fields might revolve around ongoing, established issues that do not fluctuate as dramatically on a month-to-month basis compared to more reactive topics.

The larger differences in monthly stress levels in Lifestyle and Wellness and Arts, Culture and Entertainment could be due to their close connection with personal and societal well-being, which has been heavily influenced by the pandemic's impact on daily life and cultural engagement.

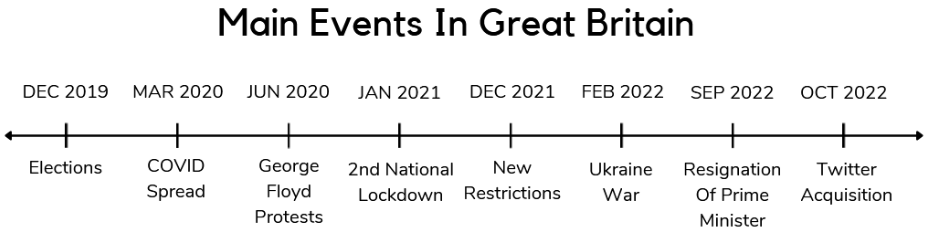
The stress in political discussions is also notably variable, likely reflecting the direct impact of political developments, elections, and legislative actions which tend to generate immediate and strong public responses.

Despite the pandemic's profound impact on the sectors of Business and Finance, Science and Technology and Health and Wellness, the relative consistency in stress scores may indicate a more measured, continuous engagement with these topics, where the discussions are less about immediate reactions and more focused on long-term impacts and solutions.

5.2 Great Britain Analysis

The Great Britain Analysis provides an in-depth exploration of the pandemic's impact on public sentiment and policy responses in the UK. It begins with an overview of key events, followed by detailed emotional analyses through T-tests and fixed effect regressions. Further, it assesses the influence of COVID-19 data and public health restrictions. This section aims to highlight how specific events and policies shaped public reactions and adapted to the evolving challenges throughout the pandemic.

5.2.1 Main Events in Great Brittain



The timeline provided captures the critical events in Great Britain that span from political shifts to global and national crises. The series begins with key political developments and progresses through the initial and severe phases of the COVID-19 pandemic, societal responses to global social justice movements, shifts due to international conflicts, and notable political changes. Each of these moments has played a role in shaping the discourse, policies, and emotional responses within the UK, which is important for the analysis.

5.2.2 Emotion Analysis

The emotion analysis for Great Britain delves into shifts in public sentiment using T-tests, fixed effect regressions, and temporal changes in emotions. These methodologies assess the significance of emotional reactions to major events, accounting for individual variability and illustrating trends over time. This approach offers an understanding of how significant occurrences influence collective emotions in the UK.

5.2.2.1 T-Tests

The t-test analysis for Great Britain specifically targeted changes in emotional expressions on Twitter before and after the COVID-19 pandemic hit. This method quantified differences by comparing mean scores of various emotions across two periods, pre-pandemic, and post-pandemic.

T-Test Results

emotion	before_mean	after_mean	t_statistic	p_value
fear	0.036	0.040	-8.565	0.000
anger_y	0.168	0.182	-11.822	0.000
joy_y	0.437	0.409	16.449	0.000
optimism	0.239	0.244	-4.170	0.000
sadness_y	0.157	0.166	-8.969	0.000
sentiment_y	0.295	0.247	12.325	0.000
stress_score	0.001	0.001	-0.521	0.602
offensive_score	0.106	0.109	-5.455	0.000
hate_score	0.046	0.045	5.539	0.000

Here are the detailed findings from the analysis:

- **Fear:** There was a slight increase in the expression of fear, from a mean of 0.036 before the pandemic to 0.040 afterward, with a t-value of -8.565 indicating this change is statistically significant ($p < 0.001$).
- **Anger:** Anger also showed a statistically significant increase, rising from 0.168 to 0.182 ($t = -11.822$, $p < 0.001$), reflecting growing frustration or irritation among the public.
- **Joy:** Joyful expressions decreased markedly from 0.437 to 0.409 ($t = 16.449$, $p < 0.001$), suggesting a reduction in positive sentiments as the pandemic progressed.
- **Optimism:** A subtle decrease in optimism was observed, with mean scores slightly dropping from 0.239 to 0.244 ($t = -4.170$, $p < 0.001$), indicating a cautious or pessimistic outlook among Twitter users in the UK.
- **Sadness:** Sadness increased, with mean levels rising from 0.157 to 0.166 ($t = -8.969$, $p < 0.001$), highlighting an overall increase in feelings of sorrow and distress.
- **Overall Sentiment:** There was a significant drop in overall sentiment, from 0.295 to 0.247 ($t = 12.325$, $p < 0.001$), suggesting a general downturn in mood.

- **Stress:** The change in stress levels was not statistically significant, with the t-statistic of -0.521 and a p-value of 0.602, indicating stable stress levels despite the pandemic.
- **Offensive Content:** There was a minor but significant increase in offensive content, from 0.106 to 0.109 ($t = -5.455$, $p < 0.001$), possibly due to heightened tensions and polarized sentiments.
- **Hate:** Interestingly, the expression of hate slightly decreased, contrary to the increase in anger and offensive content, with means shifting from 0.046 to 0.045 ($t = 5.539$, $p < 0.001$).

These results indicate that while there are some increases in negative emotions such as anger and sadness, the overall complexity of emotional responses, including the decrease in hate, suggests a subtle impact of the pandemic on public mood in Great Britain.

5.2.2.2 Fixed Effect Regressions

The Fixed Effects Regression analysis conducted for the UK evaluated how individual Twitter users' expressions of various emotions changed over time in response to the COVID-19 pandemic. This analysis controlled for both observed and unobserved individual-specific variations, ensuring the changes observed are because of the pandemic's impact.

Regression Results

sentiment	coefficient	t_value	p_value	significance
fear	0.004	8.131	0.000	***
anger	0.017	15.457	0.000	***
joy	-0.030	-19.571	0.000	***
optimism	0.002	2.161	0.031	*
sadness	0.010	10.042	0.000	***
sentiment	-0.058	-15.687	0.000	***
stress score	0.000	0.093	0.926	

offensive score	0.005	9.248	0.000	***
hate score	0.000	0.008	0.994	

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

Summary of findings from 15,509 different Twitter users:

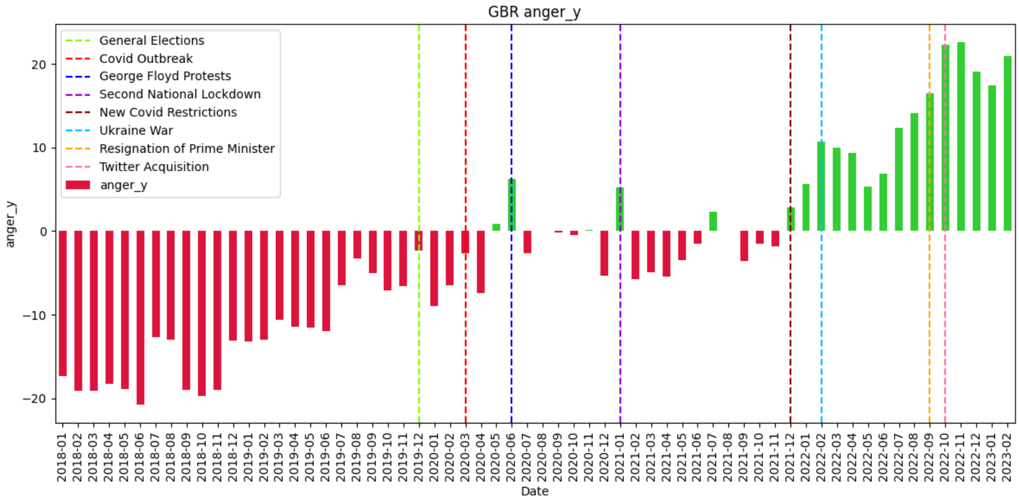
- **Fear** and **Anger:** Both saw significant increases, with coefficients of 0.004 and 0.017 respectively, indicating heightened anxiety and irritation.
- **Joy:** Experienced a significant drop, coefficient of -0.030, reflecting a decline in positive emotional expressions.
- **Optimism:** Showed a marginal but statistically significant increase, perhaps reflecting a cautious hope.
- **Sadness:** Increased notably, with a coefficient of 0.010, suggesting a rise in distress.
- **Sentiment:** Marked decrease, coefficient of -0.058, showing a significant turn towards negative overall sentiment.
- **Stress Score:** and **Hate Score:** Changes were not statistically significant, indicating stable levels of stress and hate through the pandemic.
- **Offensive Content:** Increased slightly, coefficient of 0.005, which could be linked to rising tensions.

These regression results from the UK's data indicate clear shifts in the emotional landscape of Twitter users due to the pandemic, with significant increases in negative emotions and a notable decrease in positive sentiment.

5.2.2.3 Emotion Change Over Time

The emotional analysis for Great Britain during the COVID-19 pandemic provides an overview of how significant events influenced public sentiment. By examining fluctuations in various emotions, this section identifies patterns and reactions to key events like elections, lockdowns, and political changes.

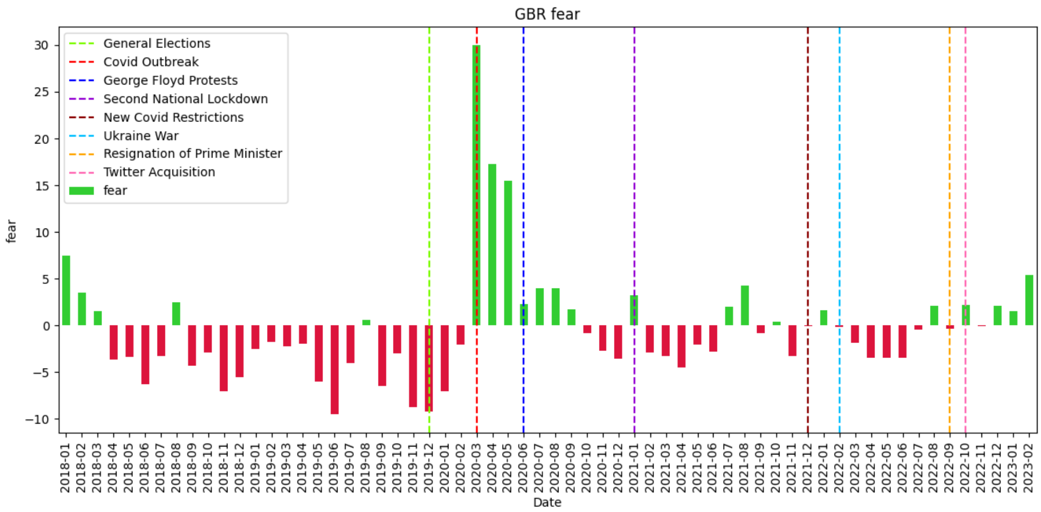
Anger



There is a notable spike in anger on June 2020, which correlates with the intense public reaction to the global Black Lives Matter protests sparked by George Floyd's death, including actions like the toppling of statues in the UK [43] and debates surrounding Brexit [44].

In early 2022, further increases in anger are observable, primarily triggered by Russia's invasion of Ukraine, illustrating the public's strong reactions to international conflicts and their implications for global stability. The latter part of 2022 shows even more pronounced spikes in anger, coinciding with the Prime Minister’s resignation [45] and Elon Musk's acquisition of Twitter, which likely fueled debates and controversies around freedom of speech on the platform.

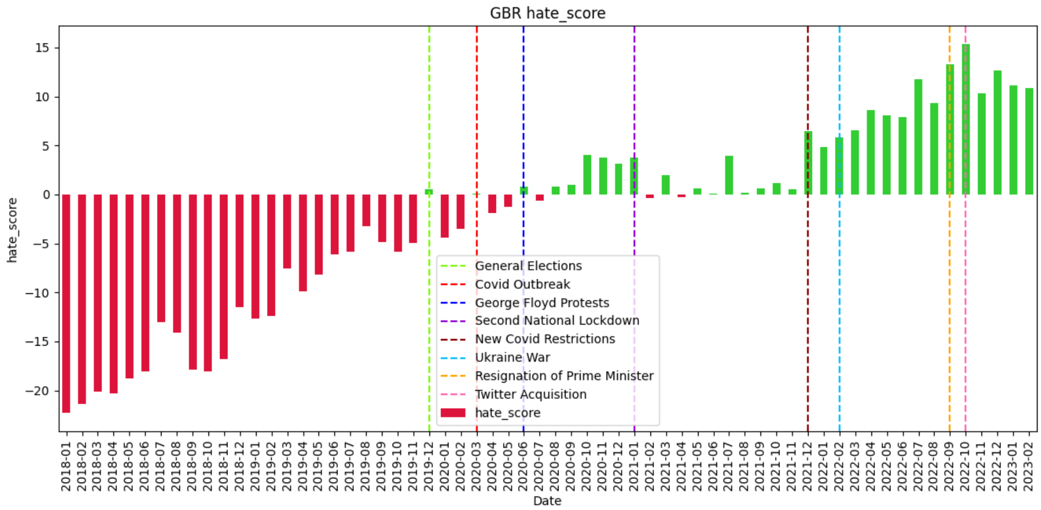
Fear



Fear levels in the UK exhibited variations aligned with significant events. A substantial spike occurred at the beginning of the COVID-19 pandemic, reflecting widespread anxiety. January 2021 saw another peak corresponding to the announcement of a second, stricter national lockdown as COVID-19 cases rose sharply [46].

The trend in fear decreased subsequently, until a resurgence in July 2021, likely due to concerns over the Delta variant's spread. The latest notable spike in fear is in February 2023, which might be linked to new, impactful events.

Hate Score

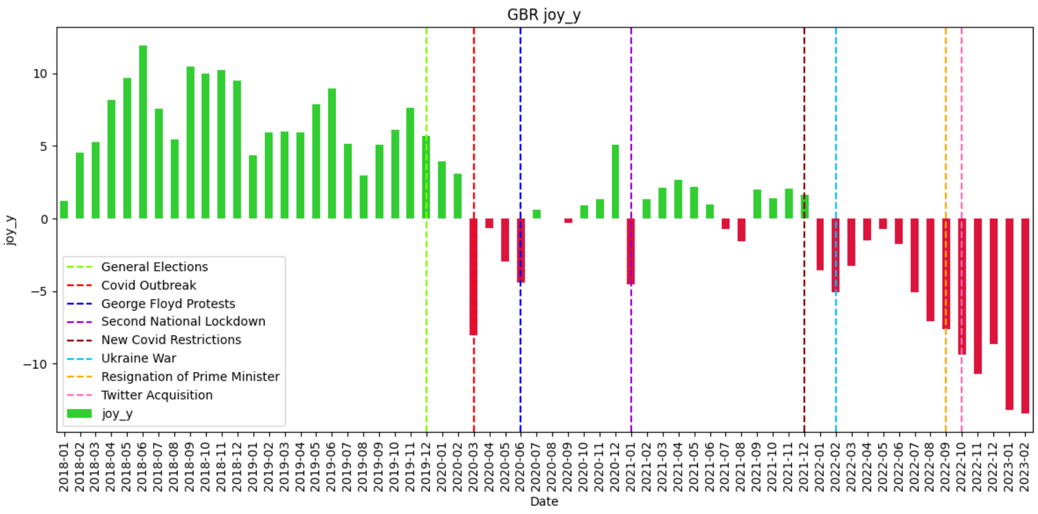


Hate score showed a gradual increase leading up to the COVID-19 pandemic, with a more pronounced rise following the onset of the pandemic.

Notable spikes are observed in June 2021, corresponding with heightened tensions during the Black Lives Matter protests and ongoing Brexit negotiations. Another significant increase occurred in December 2021, triggered by the announcement of 'Plan B' COVID-19 restrictions by Prime Minister Boris Johnson [47].

The trend continued upward with the onset of the Ukraine war and saw further escalation following the resignation of the Prime Minister and acquisition of Twitter in September/October 2022.

Joy

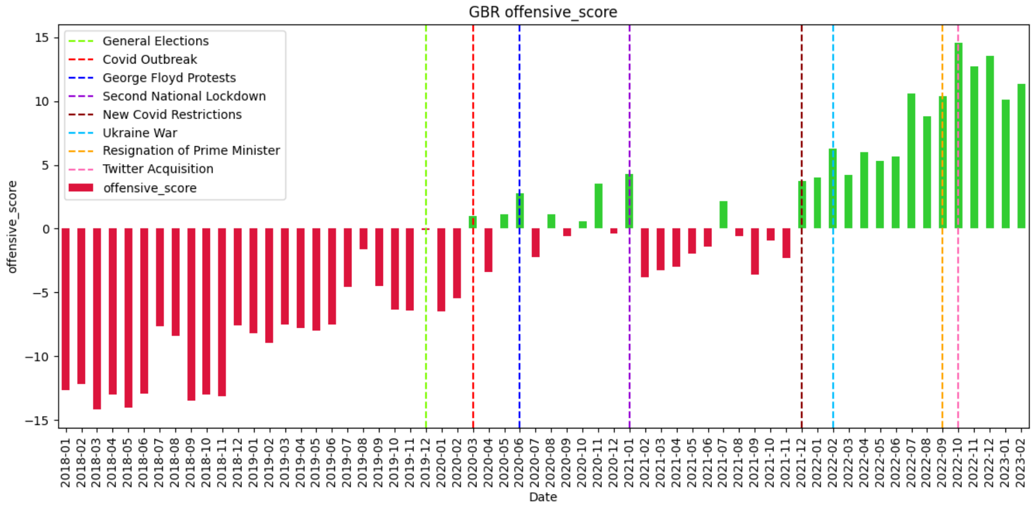


Joy levels in the UK initially plummeted with the onset of the COVID-19 pandemic, reflecting the widespread distress.

Subsequent events like the Black Lives Matter protests and developments in the pandemic forcing a second lockdown contributed to additional decreases in joy.

Notably, the trend continued to decline sharply following the onset of the Ukraine war and further during the resignation of the Prime Minister and acquisition of Twitter. Throughout this graph, joy levels show a clear sensitivity to global and national crises.

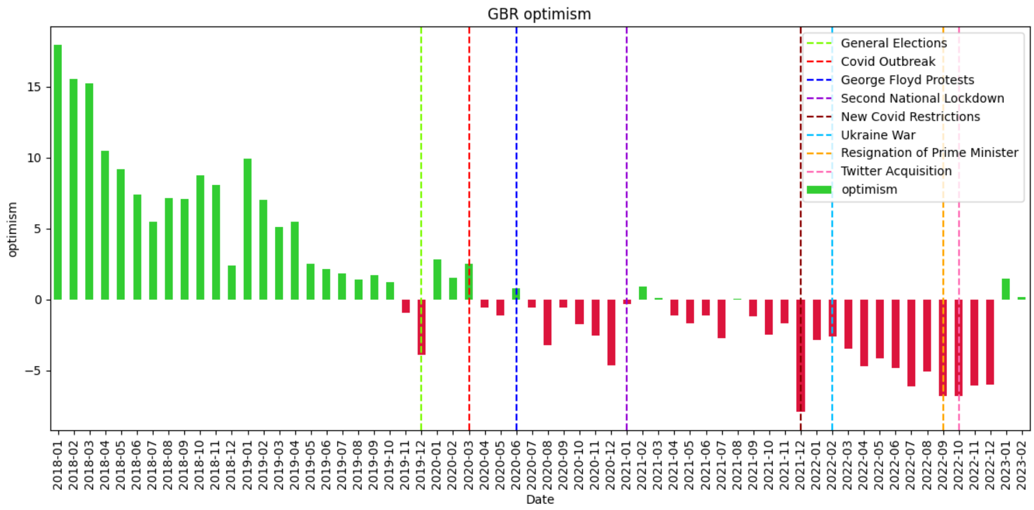
Offensive Score



The offensiveness score demonstrates a less direct correlation with the start of COVID-19 compared to other emotions. However, specific events such as the George Floyd protests in June 2020 and the announcement of the second national lockdown in January 2021 show noticeable increases in offensiveness.

Spikes in offensiveness occur again with the imposition of new covid restrictions, and the onset of the Ukraine war. But the largest spikes happened because of the resignation of the Prime minister and the acquisition of Twitter, suggesting these events had a significant impact on the public. The trend reflects a heightened sensitivity to global conflicts and major corporate changes.

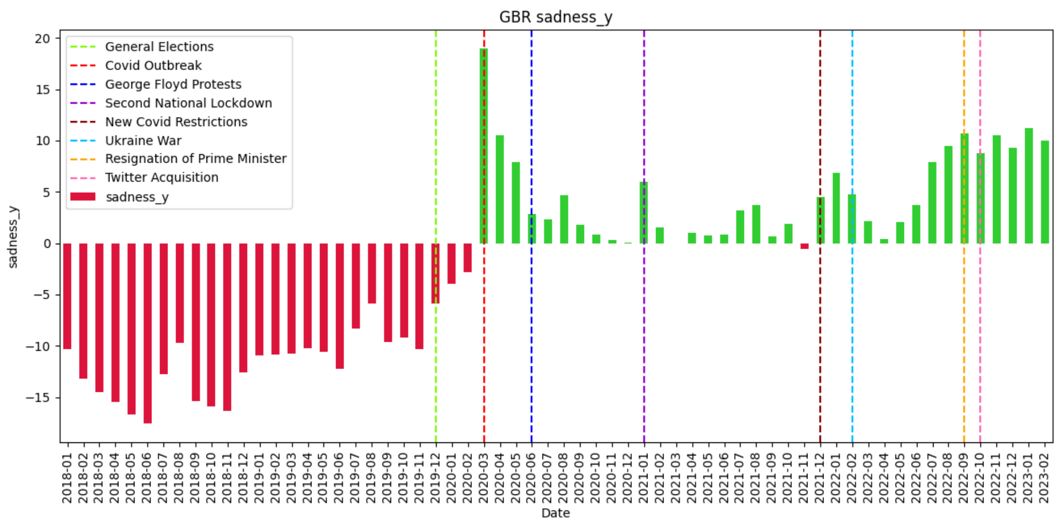
Optimism



Optimism saw its first significant decline in December 2019, influenced by the general elections [48] and severe weather conditions [49]. A brief recovery occurred early in 2020, caused by the resolution of Brexit, which provided some closure and certainty.

However, this rebound was short as the COVID-19 pandemic began, leading to prolonged periods of negative optimism levels. The most substantial decline occurred in December 2021, coinciding with a surge in COVID-19 cases and the imposition of additional restrictions by the government. This period marks the lowest point of optimism on record during the timeframe observed, with optimism remaining negative after the following events.

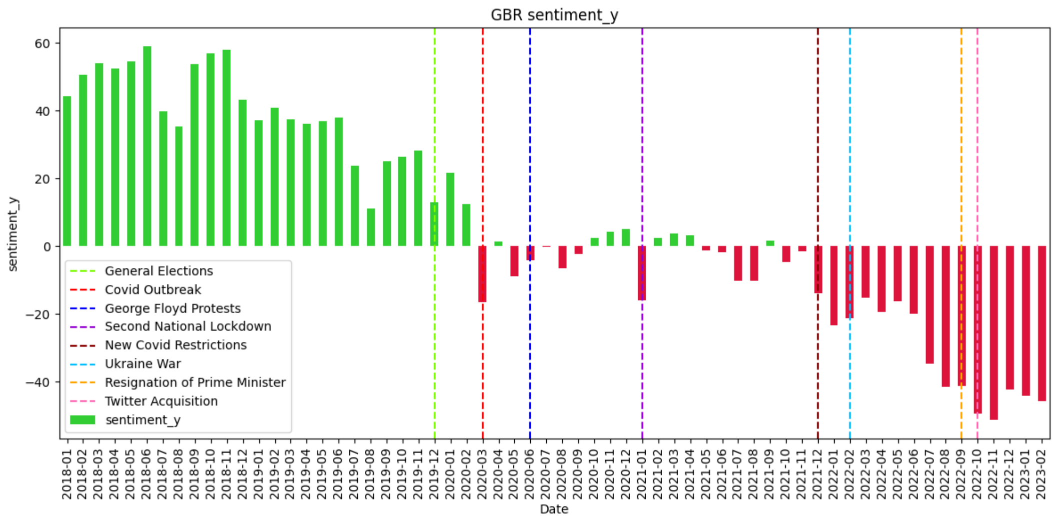
Sadness



Before the pandemic, sadness levels in the UK were generally negative, indicating lower levels of this emotion. With the outbreak of COVID-19, sadness sharply increased, reaching its highest point at the pandemic's onset.

The positive trend remained persistent throughout the whole timeline, with bigger spikes during significant incidents such as the George Floyd protests, various phases of COVID-19 lockdowns, and the outbreak of the resignation of the Prime Minister. Each event seems to have contributed to notable increases in sadness, indicating a direct correlation between them and heightened emotional responses.

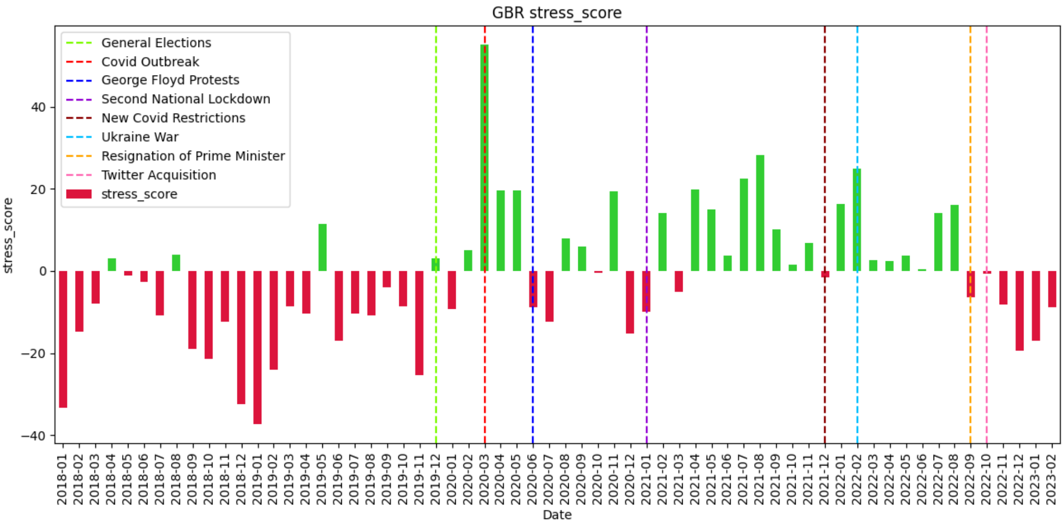
Sentiment



Overall sentiment in the UK had a sharp decline with the start of the pandemic. This trend of lower sentiment persisted, with January 2021 marking a significant drop as the government announced a second lockdown. The situation further deteriorated with the new COVID-19 measures and the onset of the Ukraine war towards the end of 2021 and early 2022.

Political turmoil, including the Prime Minister's resignation and the subsequent Conservative Party leadership election between July and September 2022, led to another sharp decrease in sentiment. The acquisition of Twitter in October-November 2022 also contributed to a notable negative spike, reflecting the impact of major corporate and political events on public sentiment.

Stress Score



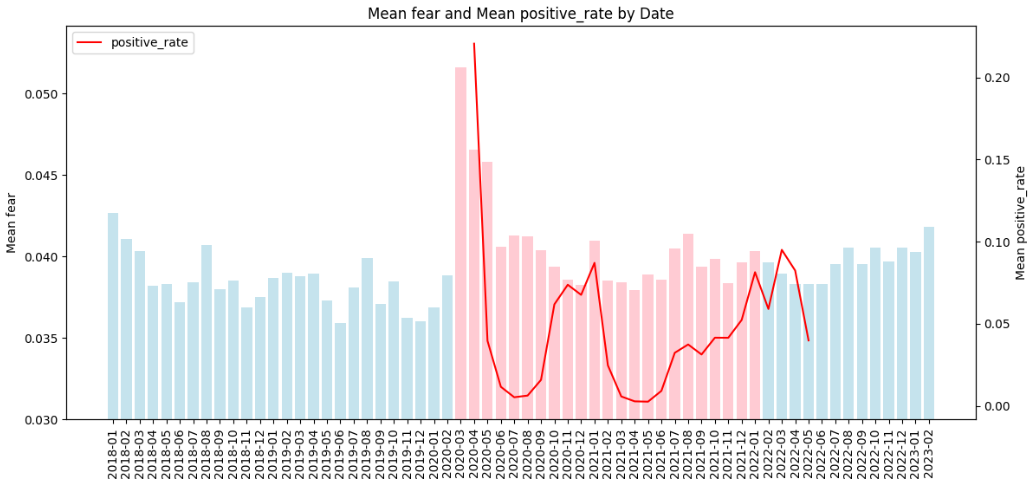
The stress chart for the UK has several peaks correlated with major events over the period observed. Initially, stress levels show significant variability but generally trend downwards until the start of the COVID-19 pandemic, which shows a big spike in stress. This peak is the highest in the series, showing the impact of the pandemic's onset on public stress levels.

The latter part of the timeline shows fluctuating but generally elevated stress levels, with notable increases around the time of the Ukraine war.

5.2.3 Covid Data Analysis

In the analysis of the COVID-19 data for Great Britain, bar plots depicting the mean levels of various sentiments were utilized alongside lines representing various COVID-19 metrics such as new cases, total cases, deaths, and vaccination rates. This allowed for a clear understanding of how pandemic-related events correlated with shifts in public emotions. Due to the extensive number of graphs generated, the focus was narrowed to the four most statistically significant graphs.

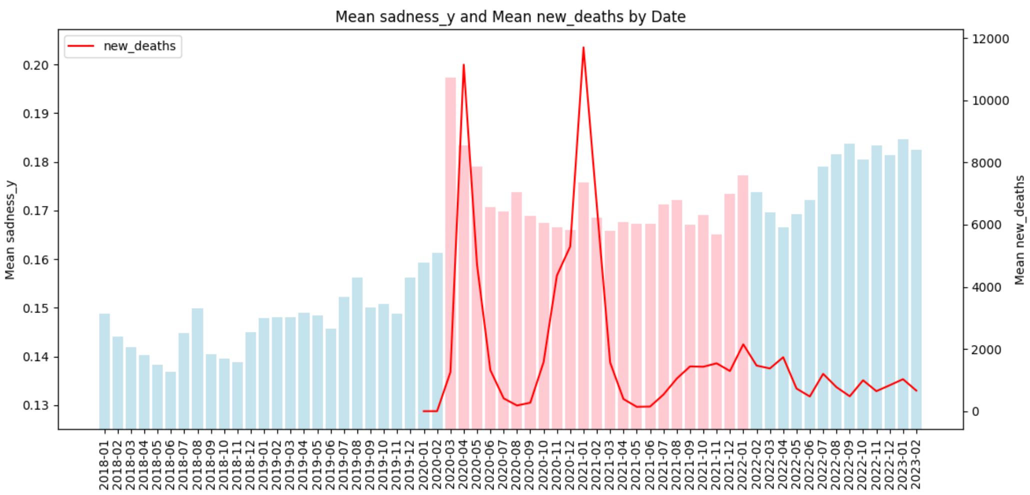
Fear and Positive Rate



Initially, as the pandemic began in early 2020, there was an increase in both the positive test rate and mean fear levels, reflecting the public's reaction to the sudden outbreak.

By early-2021, another significant spike in the positive rate was observed, which was mirrored by a peak in fear levels, suggesting a direct response to its increase. During the latter part of 2021 and into early 2022, the fluctuations in the positive rate show several peaks, with fear levels following a similar pattern, suggests that fear remained closely tied to the trends in COVID-19 data, with higher positive rates driving higher fear.

Sadness and New Deaths

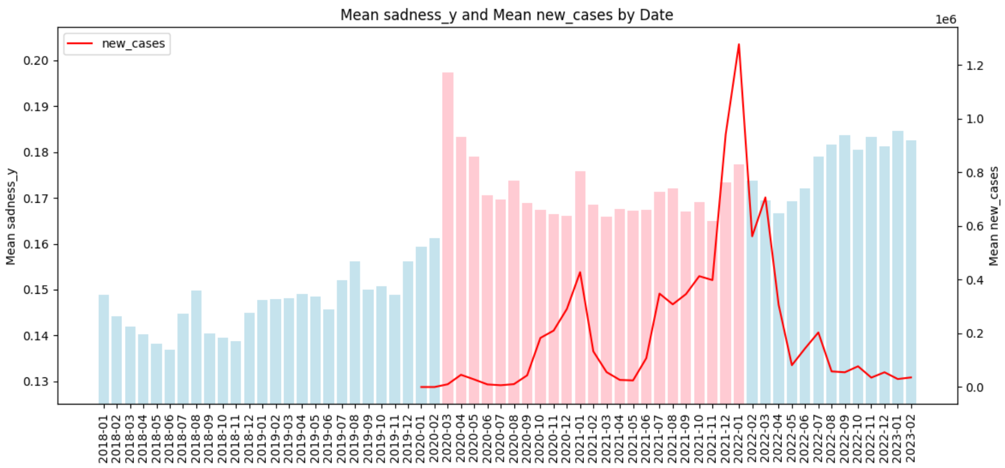


With the beginning of the pandemic in early 2020, there is a gradual increase in sadness, paralleled by a rise in new deaths, reflecting an emotional response to the escalating death toll.

The most significant peak in new deaths during early-2021 corresponds with a sharp increase in sadness, indicating intense emotional reactions to the most severe phases of the pandemic crisis.

This can also be seen towards the end of the pandemic with another rise in new deaths coinciding with a higher level of sadness among the Tweets of users from the UK.

Sadness and New Cases

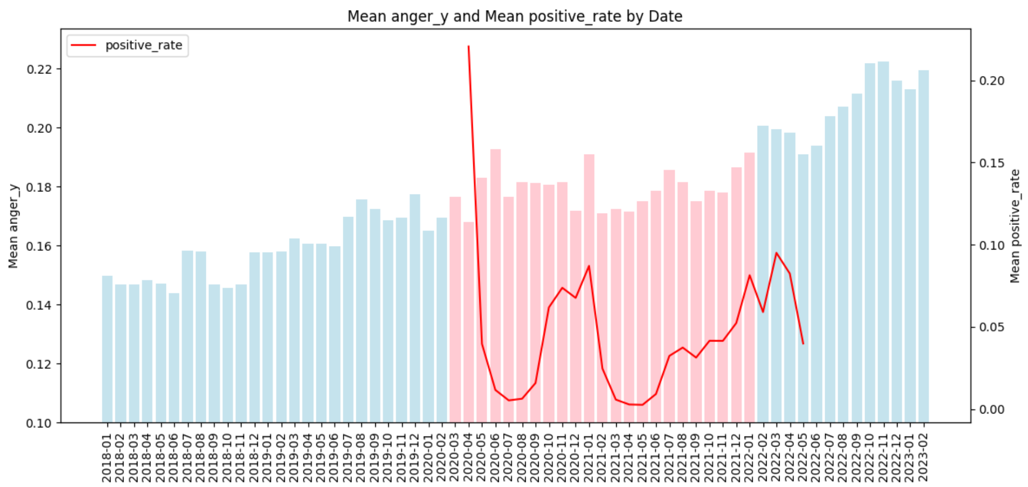


At the start of 2021, there is a noticeable spike in new cases, which corresponds to an increase in mean sadness, reflecting the public's emotional response to the resurgence of the virus.

This pattern is repeated in mid-2021, where another significant rise in COVID-19 cases aligns with a peak in sadness, indicating the ongoing impact of the pandemic on public sentiment. Again, at the start of 2022, a similar trend is observed, with a sharp increase in new cases leading to heightened sadness.

These specific points highlight how each wave of the pandemic has directly influenced the level of sadness experienced by the population, demonstrating an emotional reaction to the fluctuations in COVID-19 case numbers.

Anger and Positive Rate



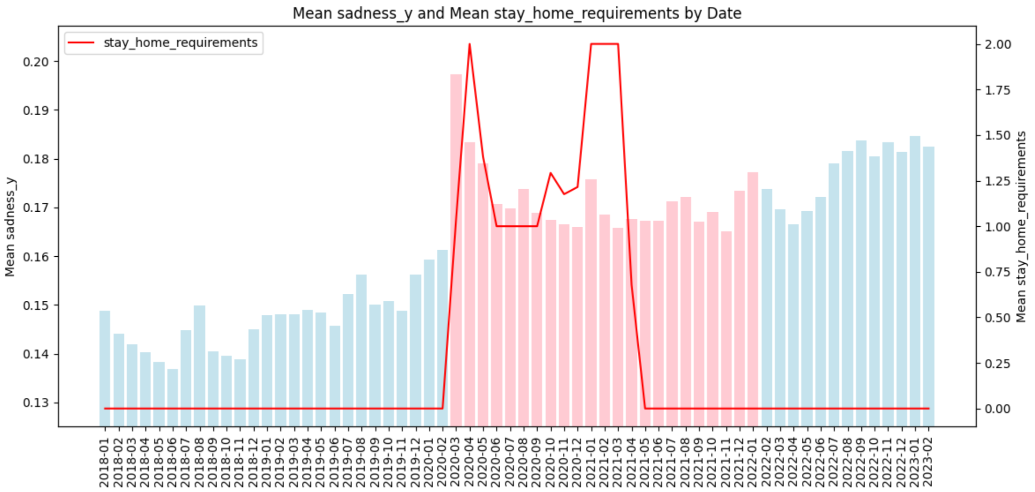
Notably, towards the end of 2020 and beginning of 2021, there is a significant rise in the positive rate, which coincides with an increase in mean anger, suggesting a heightened emotional response to a surge in COVID-19 cases during this period.

This pattern is observed again in mid-2021, and at the beginning of 2022, where further spikes in the positive rate trigger a corresponding increase in anger, reflecting ongoing public frustration or anxiety related to the persistent challenges of the pandemic.

5.2.4 Covid Restrictions Analysis

For the analysis of COVID-19 restrictions, bar charts were used again to show average public sentiment, and line graphs to detail the implementation timelines of restrictions like income support, school closures, and vaccination policies. This method provides a visualization of how different policy measures influenced public sentiment over time. Given the extensive data collected, the analysis focused on the three most statistically significant charts.

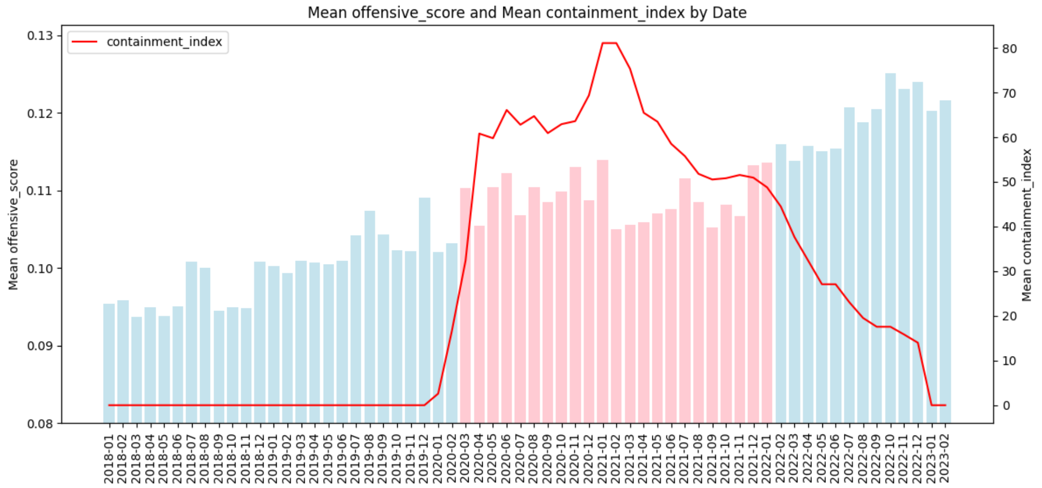
Sadness and Stay Home Requirements



There appears to be a significant correlation between increased stay-home requirements and spikes in the mean sadness level. Particularly, significant peaks in sadness can be observed during times when the stay-home requirements are at their highest. This suggests that stricter stay-home measures may correlate with higher expressions of sadness among the population.

Notably, there are sharp peaks in sadness around April 2020 and again in January 2021. These peaks closely follow the introduction or tightening of stay-home requirements, which could be a response to the initial outbreak of COVID-19 and subsequent waves of the virus.

Offensiveness and Containment Index

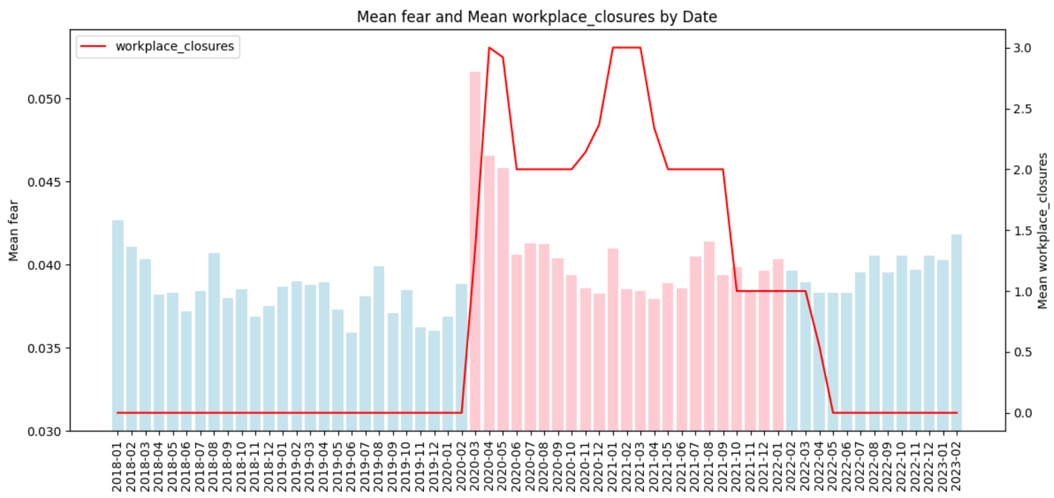


There is a noticeable correlation where increases in the containment index are followed by rises in the mean offensive score. This trend suggests that stricter COVID-19 restrictions may correlate with higher expressions of offensive content in tweets.

The offensive score peaks notably in mid-2020 and early 2021, which aligns with the period when the containment index reaches its highest points. This period likely corresponds to the strictest lockdown measures, indicating that heightened restrictions may have led to increased frustrations or discourse among the public.

After the initial peak, as the containment index begins to decline, the mean offensive score also shows a general decrease.

Fear and Workspace Closures



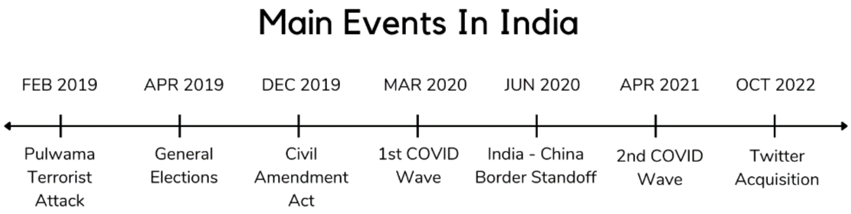
A significant increase in fear expression can be seen coinciding with the initial rise in workplace closures. This spike in fear likely reflects concerns about job security, economic impact, and the health implications of the pandemic.

The highest peaks in workplace closures align with elevated fear levels, with the second highest peak occurring around January 2021, which seems to have caused another surge in fear.

5.3 India Analysis

The India Analysis explores how the pandemic has affected public sentiment and policy responses in India. It starts with an overview of key events, followed by detailed emotional analyses using T-tests and fixed effect regressions. It also examines the impact of COVID-19 data and public health measures. This section examines how specific events and policies have shaped public reactions during the pandemic in India.

5.3.1 Main Events in India



The timeline depicts key events in India, starting with significant national incidents such as the Pulwama attack, through major political events including general elections and legislative changes, to critical health crises marked by the COVID-19 waves. These events are crucial for understanding the shifts in public sentiment and discourse within India, as they directly impact the emotional landscape of the nation.

5.3.2 Emotion Analysis

The emotion analysis for India investigates changes in public sentiment through T-tests, fixed effect regressions, and shifts in emotions. These techniques measure the significance of emotional responses to notable events, while also considering individual variations and illustrating trends over time.

5.3.2.1 T-Tests

The t-test analysis for India examines the shifts in emotional expressions on Twitter before and after the outbreak of the COVID-19 pandemic. By comparing the mean scores of various emotions from tweets collected in distinct periods, pre-pandemic and post-pandemic, the analysis quantifies the changes caused by the pandemic.

T-Test Results

emotion	before_mean	after_mean	t_statistic	p_value
fear	0.036	0.043	-11.126	0.000
anger_y	0.155	0.185	-19.361	0.000

joy_y	0.398	0.365	13.958	0.000
optimism	0.308	0.290	9.366	0.000
sadness_y	0.139	0.159	-15.349	0.000
sentiment_y	0.366	0.270	17.896	0.000
stress_score	0.001	0.001	-3.030	0.002
offensive_score	0.093	0.103	-16.800	0.000
hate_score	0.036	0.040	-12.939	0.000

Findings from the Analysis:

- **Fear:** Fear levels experienced a slight increase from 0.036 to 0.043, suggesting growing worry among the populace ($t = -11.126$, $p < 0.001$).
- **Anger:** There was a marked increase in anger, rising from 0.155 to 0.185 ($t = -19.361$, $p < 0.001$), indicating heightened frustration and irritation.
- **Joy:** Joyful expressions decreased significantly from 0.398 to 0.365 ($t = 13.958$, $p < 0.001$), reflecting a notable decline in positive sentiments as the pandemic unfolded.
- **Optimism:** A subtle decrease in optimism was observed, with scores dropping from 0.308 to 0.290 ($t = 9.366$, $p < 0.001$), signaling a shift towards a more cautious or pessimistic outlook among Twitter users in India.
- **Sadness:** Expressions of sadness rose from 0.139 to 0.159 ($t = -15.349$, $p < 0.001$), highlighting an overall increase in emotional distress.
- **Overall Sentiment:** The overall sentiment significantly decreased from 0.366 to 0.270 ($t = 17.896$, $p < 0.001$), indicating a general downturn in mood.
- **Stress:** While the change in stress levels was minor, it was still statistically significant ($t = -3.030$, $p = 0.002$), suggesting an uptick in anxiety or stress.

- **Offensive Content:** An increase in offensive content was noted, from 0.093 to 0.103 ($t = -16.800$, $p < 0.001$), possibly due to heightened tensions and polarized sentiments.
- **Hate:** Hate expression also rose slightly, from 0.036 to 0.040 ($t = -12.939$, $p < 0.001$), suggesting an increase in hostile sentiments during the pandemic.

These results reveal significant statistical changes in all emotions following the onset of COVID-19, mirroring the global emotional response to the pandemic but also highlighting specific trends within India.

5.3.2.2 Fixed Effect Regressions

The Fixed Effects Regression analysis for India provides an in-depth examination of how individual Twitter users' emotional expressions evolved during the COVID-19 pandemic.

Regression Results

sentiment	coefficient	t_value	p_value	significance
fear	0.006	9.305	0.000	***
anger	0.025	16.097	0.000	***
joy	-0.034	-16.523	0.000	***
optimism	-0.009	-5.494	0.000	***
sadness	0.019	13.359	0.000	***
sentiment	-0.080	-15.434	0.000	***
stress score	0.000	2.076	0.038	*
offensive score	0.010	14.918	0.000	***
hate score	0.002	6.847	0.000	***

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

Overview of findings from 9,794 unique Twitter users:

- **Fear** and **Anger:** Both emotions show marked increases, with coefficients of 0.006 and 0.025 respectively, highlighting significant growth in these feelings among Twitter users in India.

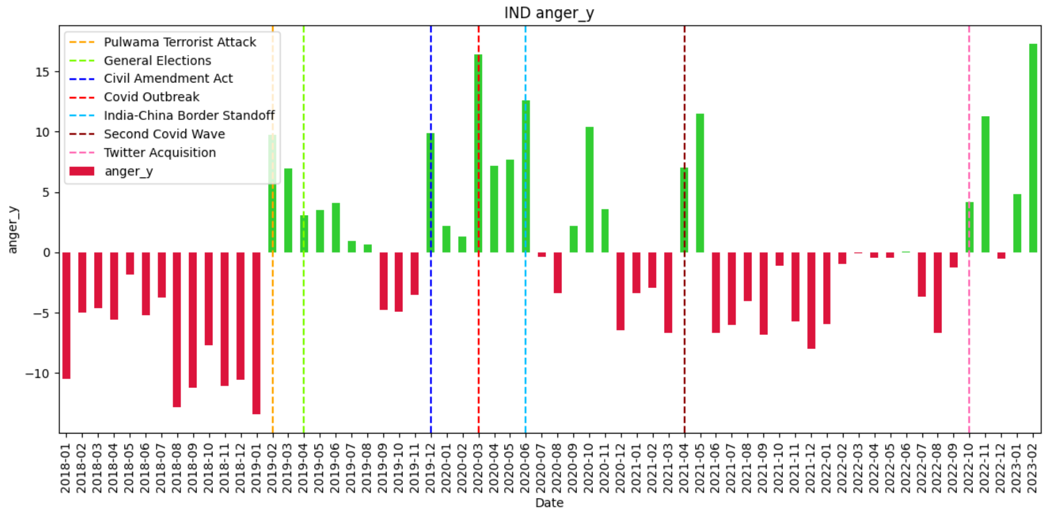
- **Joy:** There is a substantial reduction in expressions of joy, as indicated by a coefficient of -0.034, depicting a decrease in positive emotions.
- **Optimism:** A notable decline in optimism, coefficient of -0.009, suggests a shift towards a more pessimistic outlook among users.
- **Sadness:** Increased significantly, with a coefficient of 0.019, pointing to enhanced feelings of distress.
- **Sentiment:** The overall sentiment dramatically shifted towards the negative, with a significant decrease indicated by a coefficient of -0.080.
- **Stress Score:** Exhibited a minor but statistically significant increase, suggesting a slight rise in stress levels.
- **Offensive Content:** Increased, with a coefficient of 0.010, indicating a rise in aggressive or confrontational language.
- **Hate Score:** Also saw an increase, with a coefficient of 0.002, highlighting a growth in expressions of hate.

These results from the regression analysis show significant shifts in the emotions of Twitter users in India due to the pandemic, with a general increase in negative emotions and a decline in positive sentiments.

5.3.2.3 Emotion Change Over Time

The emotional analysis of India during the pandemic offers an insight into the ways significant events impacted public sentiment. This section explores the variations in different emotional responses, pinpointing patterns, and reactions to major events such as elections, COVID-19 waves, and political shifts.

Anger



Anger in India saw its first significant rise in February 2019 following the Pulwama terrorist attack [50]. This heightened state of anger was sustained through to May 2019, influenced by the general elections during that period [51].

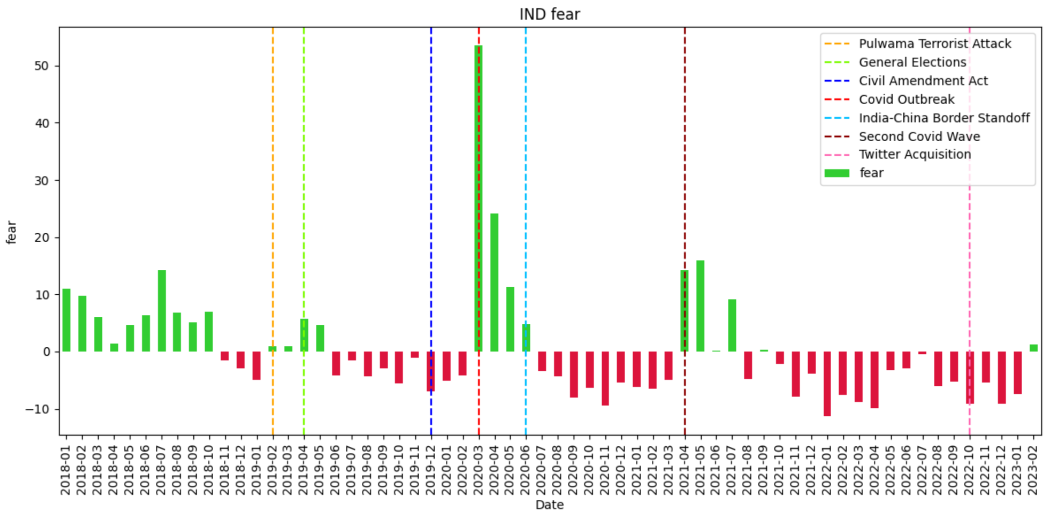
December 2019 witnessed another spike in anger due to the enactment of the Citizenship Amendment Act by the Indian Parliament [52]. This increase in anger was prolonged into the early months of the pandemic, with the pandemic causing an even bigger increase in anger.

The India-China Border Standoff in June 2020 further fueled anger across the nation [53], marking another peak in the graph.

April and May 2021 see yet another big spike, triggered by the severe second wave of COVID-19, along with vaccine shortages and distribution issues, reflecting a critical health crisis [54].

Lastly, the sale of Twitter in October and November 2022 again elevated anger levels, reflecting public reaction to major corporate events.

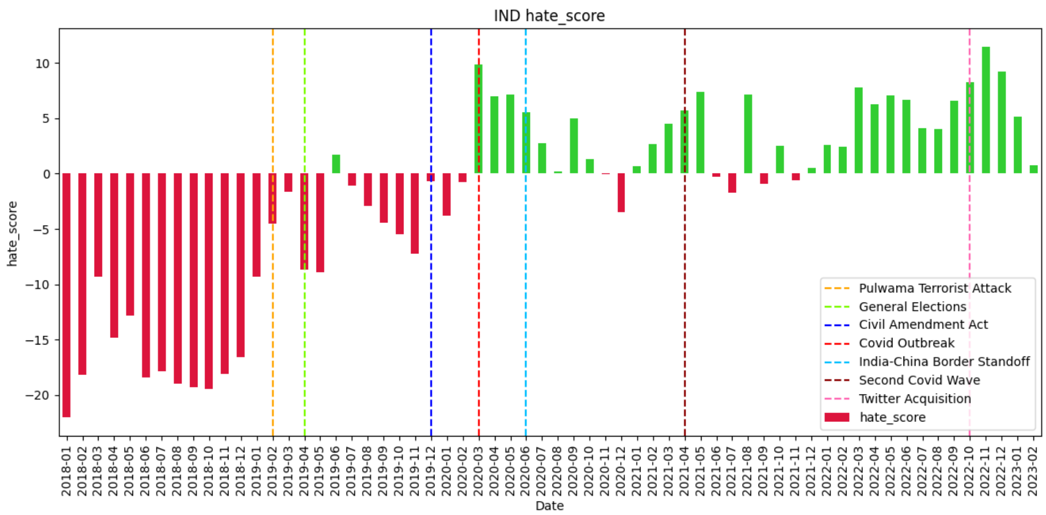
Fear



A notable surge in fear occurred during the initial months of the COVID-19 outbreak from March to June 2020. This period saw fear levels sharply increase, followed by a gradual decline in the subsequent months as initial panic subsided.

However, April 2021 marked another significant increase in fear, coinciding with the devastating second wave of the pandemic. This spike reflects the country's heightened anxiety due to escalating COVID-19 cases and the strain on healthcare systems.

Hate Score

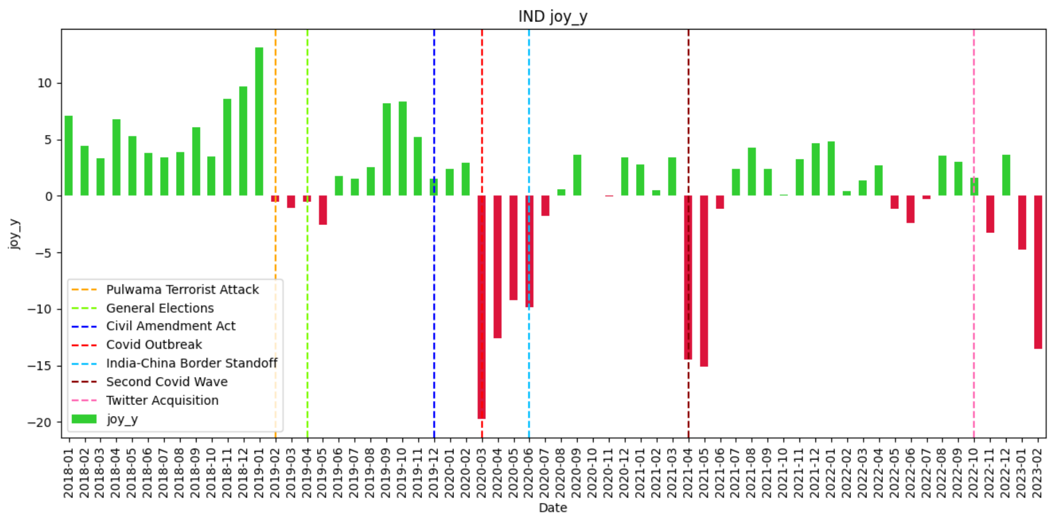


In the early months of the COVID-19 pandemic, there was a noticeable increase in hate scores in India, aligning with global uncertainty and national distress. A significant spike in hate occurred in September 2020, coinciding with the peak of daily COVID-19 cases in the country [55].

A subsequent decrease in December 2020 reflects the reduction in daily cases from the September peak, leading to a temporary decrease in negative sentiments.

However, the trend in hate scores quickly escalated again with the onset of the second wave of the pandemic, indicating renewed public distress and frustration. Later, the graph shows further increases in hate scores, which coincide with global events such as the Ukraine invasion and the Twitter acquisition.

Joy



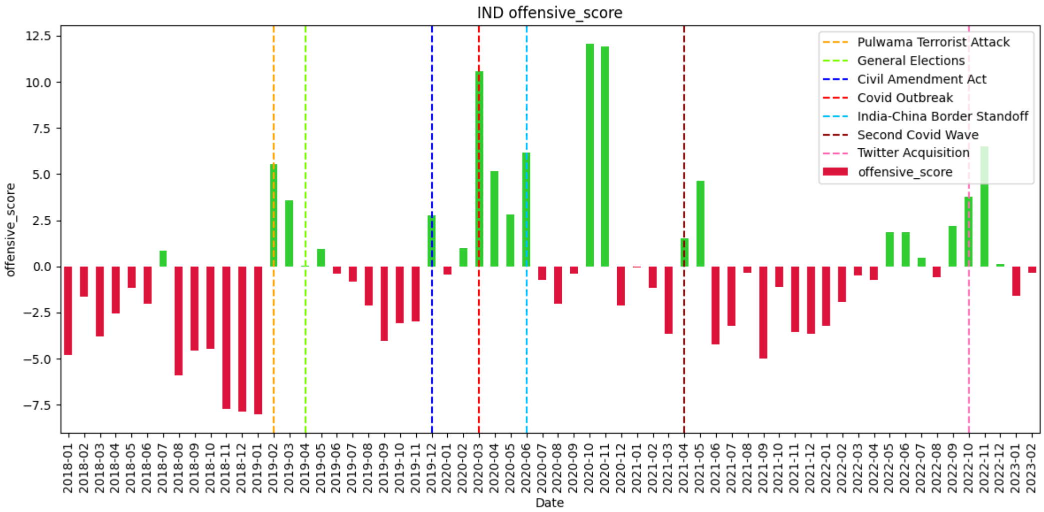
Joy experienced significant declines during key phases of the COVID-19 pandemic. The onset of the pandemic in the early months of 2020 saw a big decrease in joy, reflecting the immediate impact of the virus and the uncertainty it brought.

Another notable decline occurred in April and May of 2021, corresponding to the second wave of the pandemic.

Additionally, other events such as the Pulwama terrorist attack and the Civil

Amendment Act appear to have briefly influenced joy levels, though not to the extent seen during the pandemic's peaks.

Offensive Score

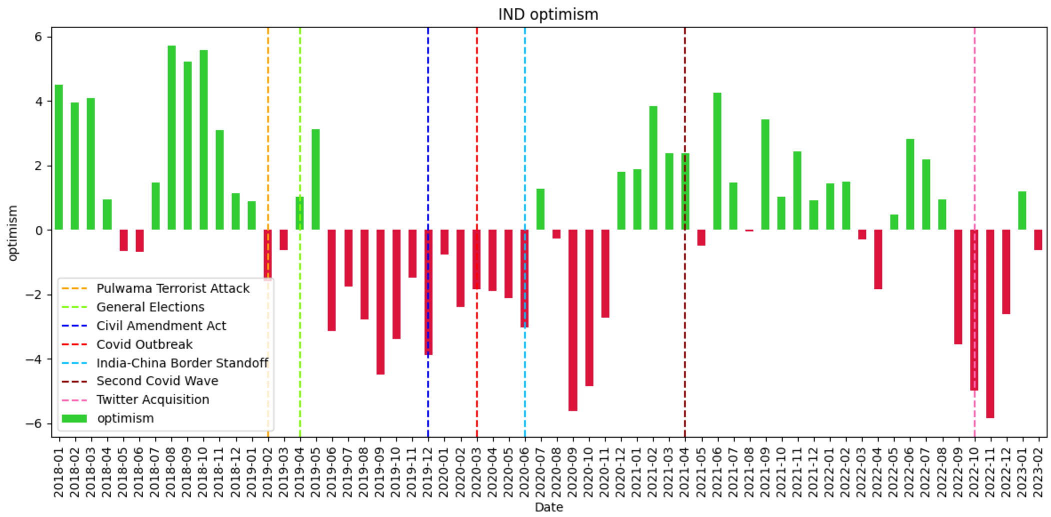


The Pulwama terrorist attack in February 2019 marked an early peak in offensiveness, reflecting the national outrage and tension following the attack. In December 2019, the passage of the Civil Amendment Act again spurred a noticeable increase in offensive content, indicating strong public reactions to political developments.

The onset of the pandemic saw a continuation of this trend, with heightened offensive scores as the country grappled with the health crisis. October and November 2020 witnessed further spikes during the intensification of the farmers' protests, which generated significant social and political debate [56].

The second wave of COVID-19 in April and May 2021 also led to an increase in offensiveness, same with the acquisition of Twitter in October and November 2022.

Optimism

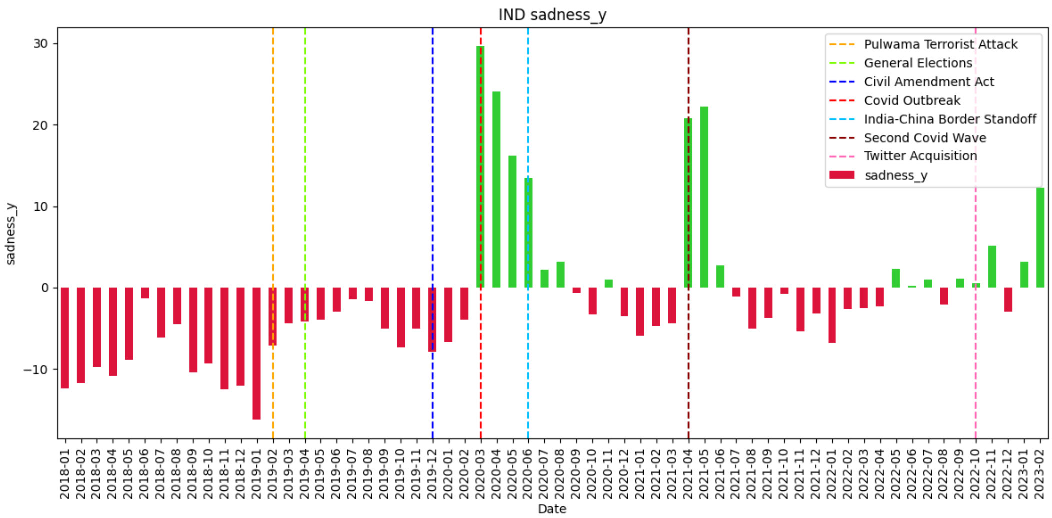


Optimism in India initially showed fluctuations demonstrated by notable peaks around certain events. After the Pulwama terrorist attack in February 2019, optimism briefly rebounded during the general elections in mid-2019, likely influenced by political hopes and aspirations.

However, the passage of the Civil Amendment Act in December 2019 led to a downturn in optimism, with the pandemic outbreak in early 2020 pushing optimism down again.

The intensification of the farmers' protests in October and November 2020, and the acquisition of Twitter in late 2022 seem to have affected optimism the most, reflected by the big negative spikes in optimism.

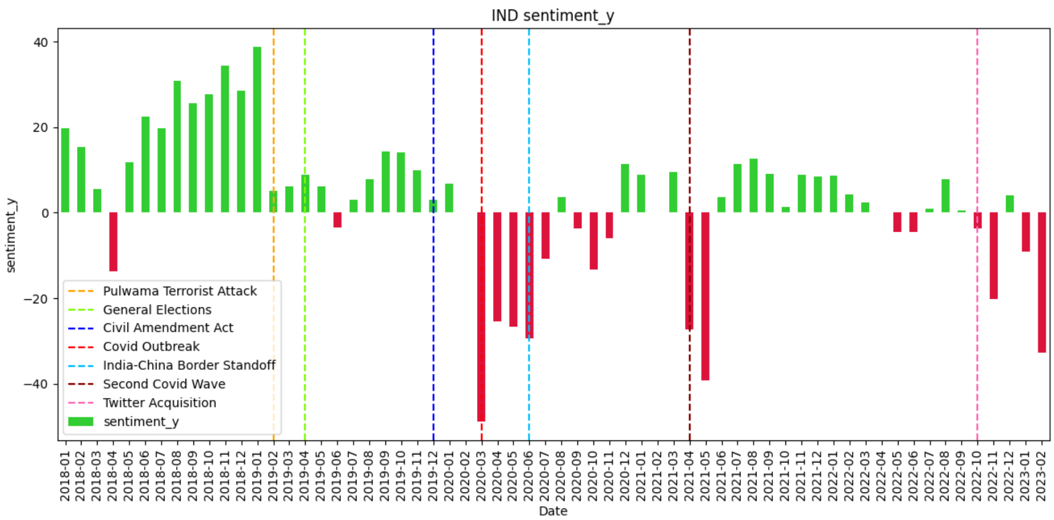
Sadness



Sadness levels show two major surges reflecting the country's response to the COVID-19 pandemic. The first surge occurred in the initial months following the outbreak in early 2020. This spike in sadness can be attributed to the sudden imposition of lockdowns, uncertainty about the disease, and the immediate economic and social disruptions that affected millions.

The second significant increase in sadness is observed during the second wave of the pandemic around April and May 2021. This period was also severe, with a dramatic rise in COVID-19 cases and deaths.

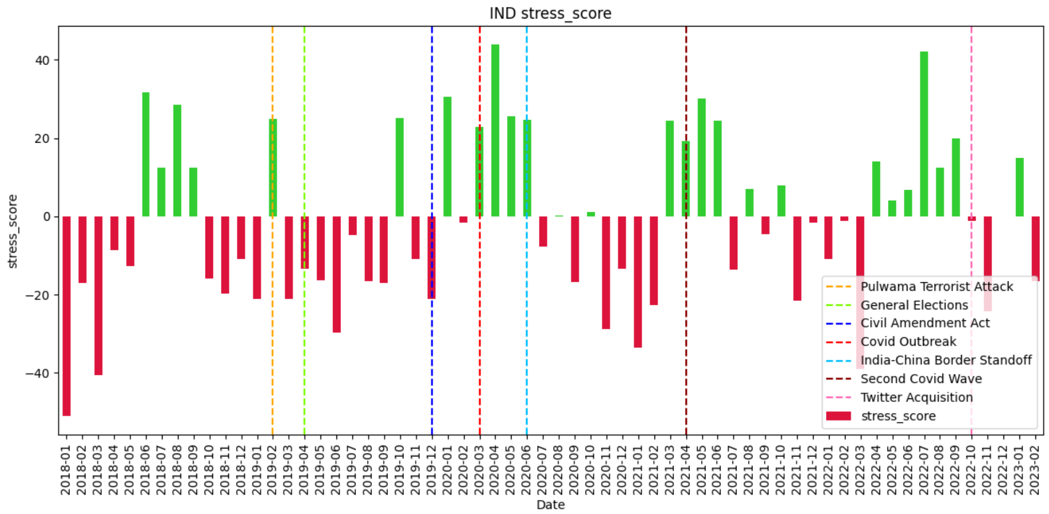
Sentiment



Overall sentiment experienced declines particularly during major COVID-19 milestones. The initial outbreak of the pandemic early in 2020 triggered a significant drop in positive sentiment, followed by further downturns during the second wave in April and May 2021, which was accompanied by a healthcare crisis.

In addition to pandemic-related drops, the graph indicates other notable decreases in sentiment around key events. The acquisition of Twitter in late 2022 also appears to have negatively impacted sentiment.

Stress Score



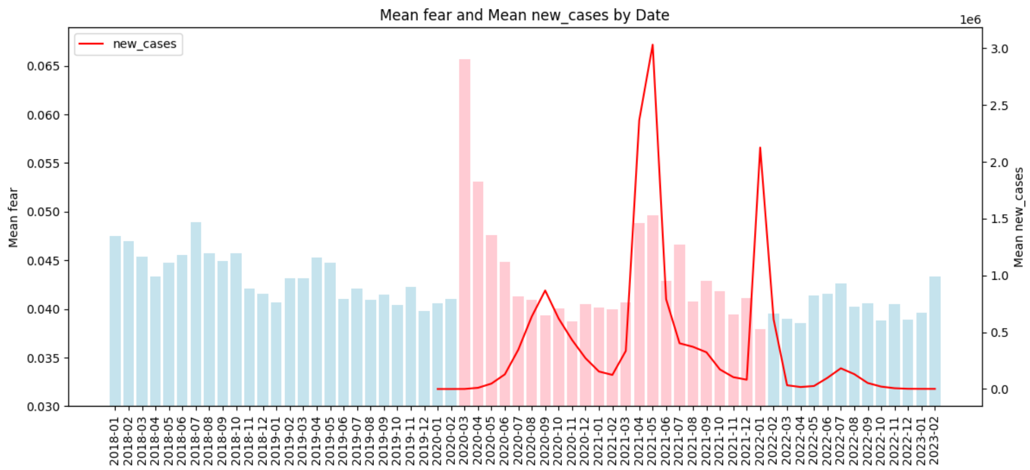
Stress levels in India demonstrate variability, with significant spikes in certain instances which might indicate data unreliability or inconsistency in how stress is measured or reported across different events.

The most definitive increases in stress are associated with three major occurrences: the Pulwama terrorist attack in February 2019, the initial outbreak of COVID-19 in early 2020, and the second COVID wave in 2021. Each of these events understandably contributed to national stress due to their severe implications on security and public health.

5.3.3 Covid Data Analysis

The study of COVID-19 data for India involved using bar charts to illustrate average sentiment levels and line graphs to display metrics like new and total cases, deaths, and vaccination rates. This approach provided a clear visualization of the relationship between pandemic events and changes in public sentiment. To manage the large volume of data effectively, the analysis was concentrated on the four graphs that showed the most significant statistical relevance.

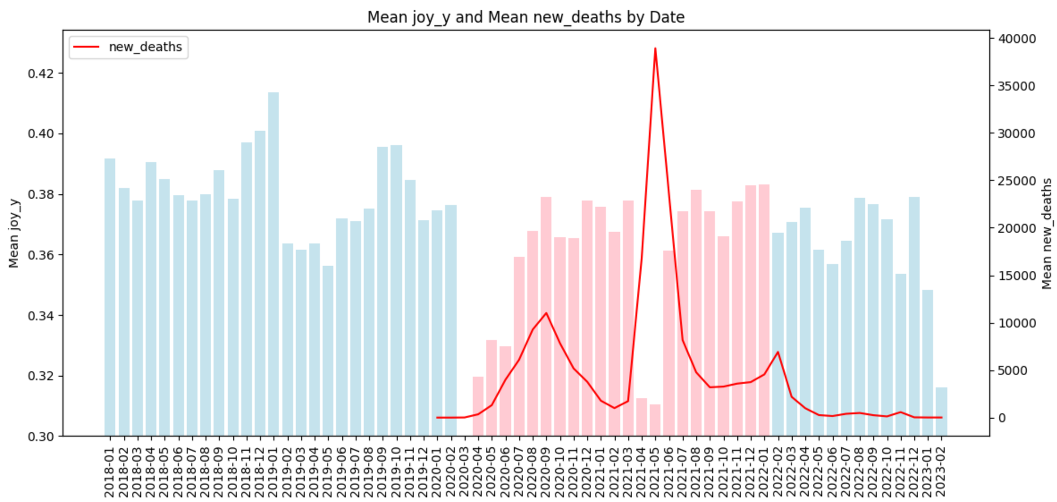
Fear and New Cases



In this graph we can see the significant impact of the second COVID-19 wave in India around April-May 2021 on public fear levels.

During this period, there is a sharp spike in new cases, which also causes a big spike in mean fear, emphasizing the direct emotional impact that the increase in new cases had on the public.

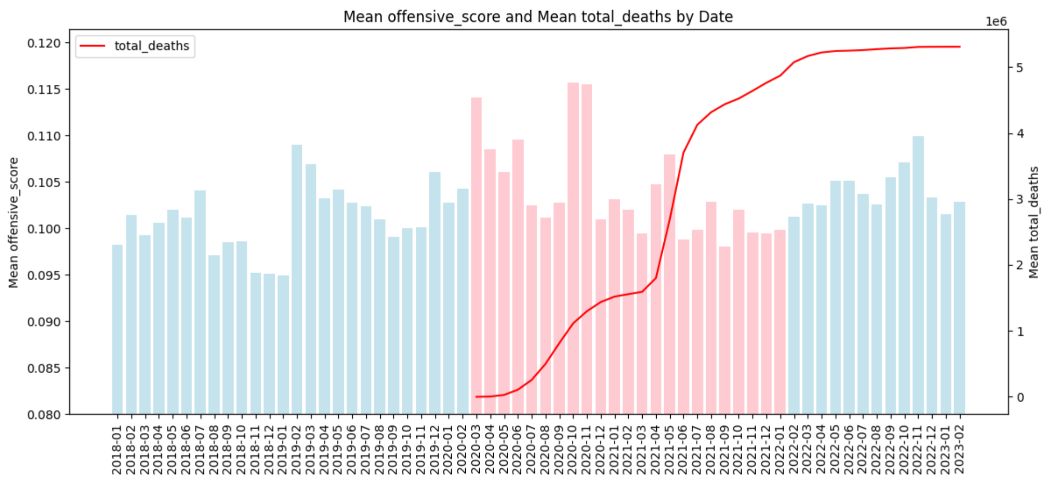
Joy and New Deaths



Initially a spike in new deaths can be observed around September 2020, which corresponds with a noticeable dip in mean joy. This trend suggests that as the number of new deaths peaked the public's sense of joy and wellbeing also declined.

Additionally, another prominent spike in deaths occurs in April-May 2021, during the second wave of COVID-19 in India, which is known for its devastating impact. This period also shows a significant decrease in joy, demonstrating once again how spikes in mortality rates directly influence public sentiment.

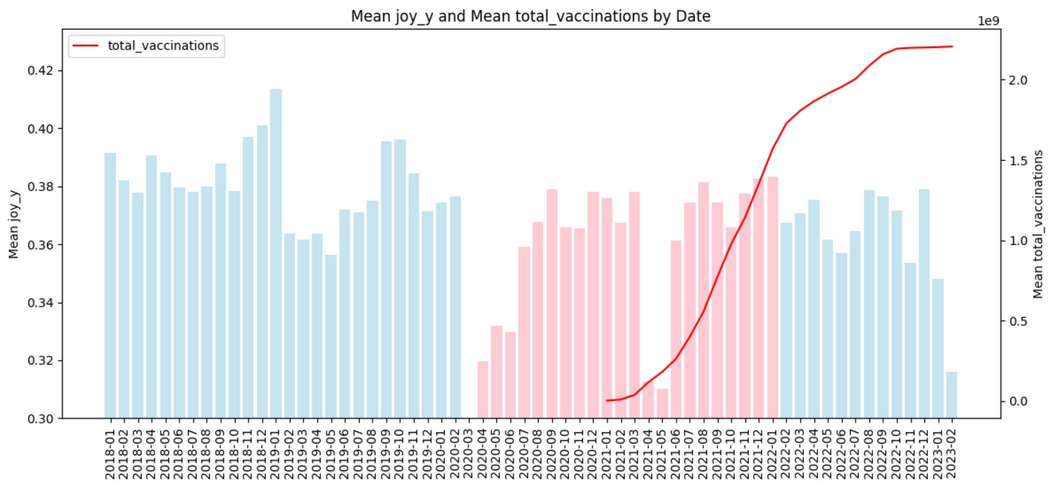
Offensiveness and Total Deaths



Toward the end of 2020, there is a noticeable rise in total deaths, which coincides with an increase in the mean offensive score. This suggests that as the death toll from the pandemic rose, there was a corresponding increase in offensive language on Twitter, reflecting possible public frustration, fear, or stress.

The graph also shows a sharp and substantial rise in total deaths during the second wave of COVID-19 in April and May of 2021, paralleled by a peak in the mean offensive score.

Joy and Total Vaccinations



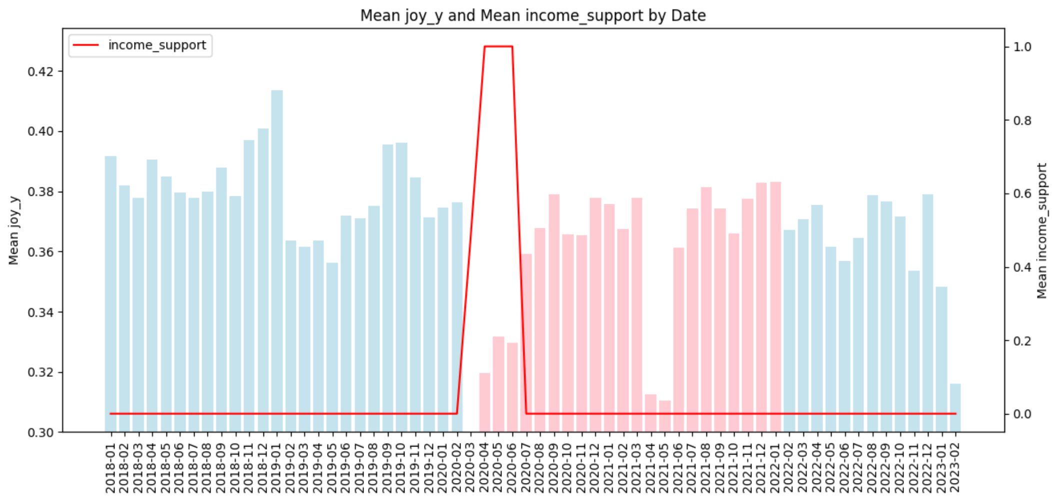
Vaccinations began in early 2021, which is a period with relatively low joy levels due to the devastating impact of the second wave of COVID-19.

As the graph progresses, it shows a significant and steady increase in total vaccinations. This increase in vaccinations coincides with a gradual increase in mean joy, suggesting that as more individuals received vaccinations, there was a corresponding rise in public joy.

5.3.4 Covid Restrictions Analysis

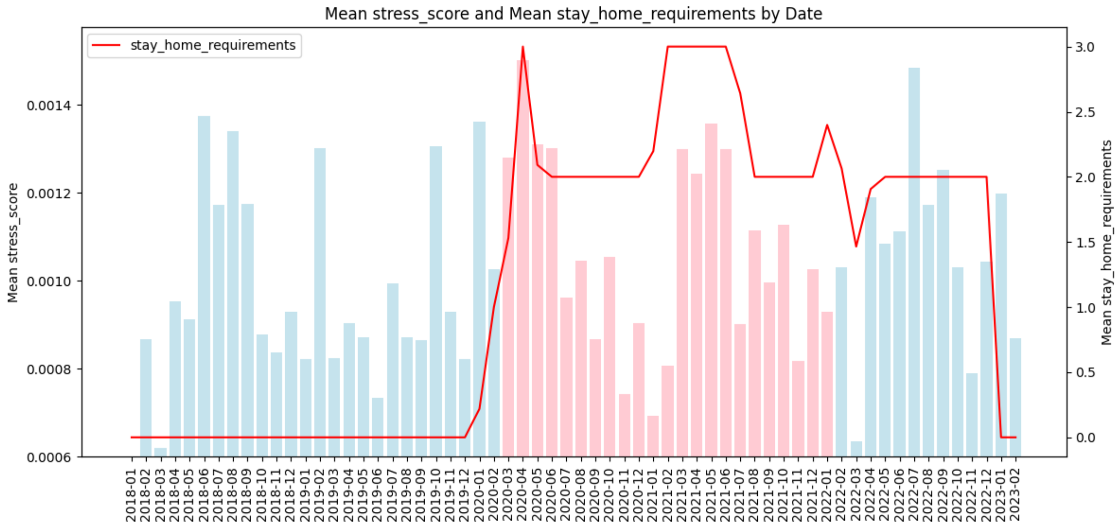
In the examination of COVID-19 restrictions, just like the data analysis part, the study employed bar charts to represent the average public sentiment and line graphs to outline the timelines for implementing various policies such as income support, school closures, and vaccination efforts. Allowing for an effective visualization of the impact of different policy measures on public sentiment throughout the pandemic. Again, due to the large amount of data gathered, the analysis concentrates on the three charts that were the most statistically significant.

Joy and Income Support



A significant peak in income support is observed in April 2020, likely reflecting the introduction of substantial financial assistance programs to mitigate the economic impact of the COVID-19 lockdowns. Interestingly, this peak does not correspond to an immediate increase in joy, which suggests that while financial support was crucial, it might not have been sufficient to significantly lift the public's spirits during the initial and intense phase of the pandemic.

Stress and Stay Home Requirements

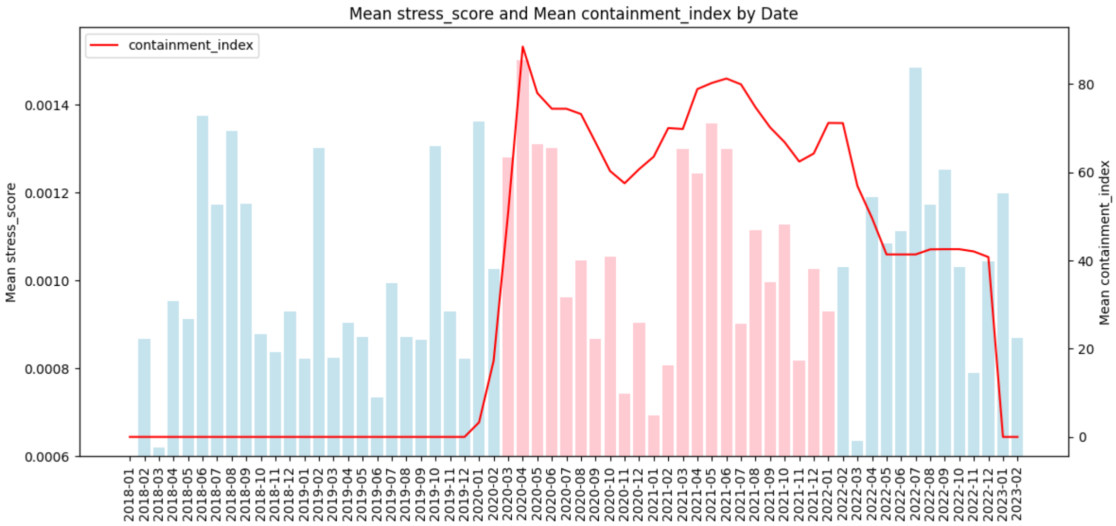


There is an initial rise in stress levels as the pandemic begins and stay-home requirements are introduced. This increase likely reflects the immediate impact of the sudden change in lifestyle.

Despite a consistent presence of stay-home requirements, stress levels show considerable variation. This suggests that while the restrictions likely contribute to stress, other factors such as the duration of the pandemic, adaptation over time, and external events also play significant roles.

As the stay-home requirements peak again in the latter part of the timeline, corresponding with the second wave in the pandemic, there is another noticeable increase in the stress levels.

Stress and Containment Index



As the containment index rises sharply in early 2020 and early 2021, corresponding to the first and second wave of the pandemic, there is an increase in stress levels. This suggests an immediate public reaction to strict measures and the uncertainty surrounding the pandemic.

Towards the later dates, as the containment index declines with the easing of restrictions, stress levels do not immediately drop but show a delayed response. This could indicate lingering effects of prolonged stress or the impact of other stressors beyond COVID-19 restrictions.

6. Limitations

This study has made substantial contributions to understanding public sentiment through Twitter analysis during the COVID-19 pandemic. However, it encountered several limitations that suggest avenues for future research:

Time Constraints

The analysis was constrained by time limitations, which restricted the depth and extent of data analysis that could be conducted. Continuous monitoring and updating of data collection and analysis methods are crucial for sentiment analysis, especially as public discourse evolves rapidly during a crisis. Future studies could benefit from longer timelines for data collection and analysis, allowing for a more comprehensive examination of sentiment trends and their evolution.

Computational Resources

Despite the application of advanced machine learning techniques and natural language processing tools, the computational resources available limited the scope of data that could be processed efficiently. This restriction impacted the ability to analyze larger datasets, which might have provided deeper insights into the nuances of sentiment shifts. Enhancing computational capacity is essential for future research, enabling the analysis of a larger dataset and potentially leading to more robust findings.

Geographical Granularity

While this research captured a broad spectrum of sentiments across different pandemic stages, it did not achieve detailed regional sentiment analysis within specific countries. The complexity and volume of the data required substantial computational power and time, which was beyond the scope of this study. Future research should focus on increasing geographical granularity by employing more localized data analyses. This would allow for a better understanding of how sentiments vary across different regions and cultural contexts.

Addressing these limitations in future studies could enhance the understanding of public sentiment, particularly in how it varies over time and across different geographic and cultural landscapes.

7. Future Research

Based on the findings and limitations of this study, several areas for further research can be proposed to enhance our understanding of public sentiment during global crises:

Enhanced Analysis of COVID Data and Restrictions

Future studies could incorporate a more detailed analysis of the correlation between public sentiment and specific COVID-19 data metrics and government-imposed restrictions. This would involve not only a broader temporal analysis but also a more segmented approach to understand how different phases of restrictions (e.g., lockdowns, mask mandates) affect public mood and sentiment. An in-depth analysis of the emotional responses to specific policy announcements or changes could offer valuable insights into public compliance and the effectiveness of different communication strategies used by governments.

Country-Specific Topic Analysis

While this thesis provided a general overview of sentiment and topics discussed globally, future research could focus on topic analysis within specific countries. This would allow researchers to capture unique cultural and regional responses to the pandemic, providing a more localized view of public sentiment. Such studies could reveal how national identity and cultural factors influence public reactions to global crises and government actions, enhancing targeted communication and intervention strategies.

Comprehensive Analysis Across All U.S. States

Given that a significant portion of the tweets analyzed originated from the USA, a detailed state-by-state analysis could be conducted. This approach would uncover state-specific sentiment trends, potentially correlating these with state-level policy decisions and pandemic impacts. It would also allow for the exploration of regional differences in sentiment within a single country, offering insights into the heterogeneity of public sentiment across different geographic and socio-political contexts.

Integration of Multimodal Data Sources

Integrating other forms of social media data, such as from Facebook, Instagram, or even non-social media sources like news outlets and blogs, could provide a more comprehensive view of public sentiment. Multimodal data analysis could uncover differences and similarities in sentiment expression across different platforms, enriching the understanding of public sentiment and its drivers.

8. Conclusion

The study presented in this thesis has made a significant contribution to understanding the dynamics of public sentiment during the COVID-19 pandemic through the analysis of Twitter data. It has highlighted how social media can serve as a powerful tool for gauging public mood and response during unprecedented global crises. The findings not only enrich academic discourse on digital sentiment analysis but also offer practical insights for policymakers, public health officials, and communication strategists who want to understand and effectively respond to public sentiment.

In response to the research questions posed in the introduction, this section articulates the key findings:

Evolution of Public Sentiment

This research delineated a clear trajectory of public sentiment over the course of the COVID-19 pandemic. It revealed that public emotions fluctuated in response to various events and announcements, reflecting a global community deeply affected by the pandemic's developments. This finding underscores the public's responsiveness to the management of the crisis and the critical role of effective communication from health authorities and governments.

Impact of Pandemic-Related Events

Significant correlations were identified between shifts in sentiment and major pandemic milestones such as lockdown initiations, vaccine announcements, and significant policy changes. These correlations highlight how events such as public health announcements and governmental policy adjustments have a profound impact on public emotions, driving sentiments from anxiety and fear to hope and relief as the situation progressed.

Geographical Variations in Sentiment

The region-specific analyses undertaken highlight the diverse ways in which different geographical areas responded to the pandemic, influenced by a combination of local policies and events, cultural factors, and the extent of the pandemic's impact. These analyses show the importance of contextual and cultural factors in shaping public sentiment and demonstrated the variability of emotional responses across different regions. By comparing sentiment across various locations, the study was able to illustrate how localized events and government actions influenced public mood and social media discourse. This approach enriches our understanding of global sentiment trends and emphasizes the need for tailored communication strategies that consider regional sensitivities and circumstances.

Role of Specific Topics in Emotional Fluctuations

The topic analysis conducted offered insights into the specific subjects that dominated Twitter discussions, revealing how different themes resonated with the public at various stages of the crisis. This analysis provided a deeper understanding of the concerns, interests, and priorities of the global community throughout the pandemic. It highlighted the power of topic-specific sentiment analysis to uncover underlying public sentiments that might not be apparent through a broader sentiment analysis. Future studies could further expand this approach by exploring inter-topic relationships and the impact of media coverage on topic popularity, providing a deeper understanding of how public attention and sentiment are directed during a crisis.

The implications of this study are many. For policymakers and health communicators, the findings stress the importance of timely, clear, and empathetic communication during crises. Public sentiment can serve as a gauge for the reception of policies and public health measures, indicating areas where additional communication or policy adjustments may be necessary. For researchers, the study highlights the utility of social media as a research tool for real-time sentiment analysis, suggesting ways for further academic exploration into methods of data collection and analysis that could enhance the robustness of sentiment analysis.

Furthermore, the study contributes to the broader discourse on the role of digital platforms in public life. As Twitter and other social media platforms continue to influence public opinion, understanding the ways in which these platforms reflect, and shape public sentiment becomes increasingly crucial. This is particularly important in the context of global crises, where digital platforms can both inform and misinform the public.

While this thesis provided substantial insights, it also encountered limitations that suggest areas for further research. The constraints of computational resources and time underscored the need for systems that can handle larger datasets and perform more complex analyses. Future research could expand upon this work by incorporating more diverse data sources and including different social media platforms.

Additionally, the geographic and cultural variations in sentiment were not fully explored in this thesis due to the scope of the study. Future research could address this gap by conducting comparative analyses of sentiment across different countries or states, particularly focusing on how local contexts influence public sentiment during crises. This could be particularly enlightening in understanding how cultural differences impact public reactions and compliance with health guidelines.

In conclusion, this thesis demonstrates the vital role of Twitter as a tool for understanding public sentiment during the COVID-19 pandemic. It can be used for future research in the realm of social media analytics, especially concerning real-time data analysis and its application to public policy and crisis management.

References

[1] Kavanaugh, A., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., Shoemaker, D. J., ... & Xie, L. (2012). Social media use by government: From the routine to the critical. Government Information Quarterly, 29(4), 480-491.

[2] Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. Journal of Medical Internet Research, 22.

[3] Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., & Chowell, G. (2021). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration. Epidemiologia, 2(3), 315-324.

[4] Medford, R., Saleh, S. N., Sumarsono, A., Perl, T., & Lehmann, C. U. (2020). An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. Open Forum Infectious Diseases, 7.

[5] Lyu, J. C., Singh, A., & Smith, T. (2020). Analysis of Twitter Sentiment During the COVID-19 Pandemic. Journal of Computational Social Science, 7(1), 145-162.

[6] Brown, D., & Zhou, Q. (2020). Impact of COVID-19 Milestone Events on Public Sentiment: A Twitter Analysis. Journal of Digital Public Health, 4(2), 58-75.

[7] Waters, J., Nicolaou, N., Stefanidis, D., Efstathiades, H., Pallis, G., & Dikaiakos, M. D. (2021). Exploring the sentiment of entrepreneurs on Twitter. PLOS ONE, 16(7), e0254337.

[8] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135

[9] Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health and Surveillance, 6(2),

e19273.

[10] Perkel, J. M. (2021). How JupyterLab has revolutionized data science workflows. Nature, 590(7845), 345-346.

[11] Spataro, W., Trunfio, G., & Sirakoulis, G. (2017). High performance computing in modelling and simulation. The International Journal of High Performance Computing Applications, 31(1), 117-118.

[12] VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc.

[13] Murrell, P. (2018). R Graphics Cookbook: Practical Recipes for Visualizing Data. O'Reilly Media, Inc.

[14] https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[15] Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., ... & Webster, S. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour, 5(4), 529-538.

[16] https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020

[17] Giustini, A., Ribeiro, M. R., Cohen, M. A., & Camargo, C. P. (2020). Impact of the COVID-19 pandemic on mental health: Realities and perspectives. International Journal of Social Psychiatry, 66(8), 810-812.

[18] Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. Nature Reviews Drug Discovery, 19(5), 305-306.

[19] https://huggingface.co/cardiffnlp/twitter-roberta-base-emotion

[20] Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. arXiv preprint arXiv:2010.12421.

[21] https://huggingface.co/bhadresh-savani/roberta-base-emotion

[22] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[23] https://github.com/allenai/allennlp-models

[24] https://huggingface.co/cardiffnlp/twitter-roberta-base-hate

[25] https://huggingface.co/cardiffnlp/twitter-roberta-base-offensive

[26] https://github.com/sanja7s/MED-DL

[27] Scepanovic, S., Martin-Lopez, E., Quercia, D., & Baykaner, K. (2020, April). Extracting medical entities from social media. In Proceedings of the ACM conference on health, inference, and learning (pp. 170-181).

[28] https://github.com/owid/covid-19-data/tree/master/public/data

[29] https://ourworldindata.org/policy-responses-covid

[30] Bowker, G. C., & Star, S. L. (2000). Sorting Things Out: Classification and Its Consequences. MIT Press.

[31] Howell, D. C. (2012). Statistical methods for psychology (8th ed.). Wadsworth, Cengage Learning.

[32] Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications. Cambridge University Press.

[33] https://www.nytimes.com/2020/06/28/world/coronavirus-updates.html

[34] Xie, Y., & Zhou, X. (2021). "The impact of the George Floyd protests on public sentiment and social change." Journal of Social Issues, 77(1), 104-121.

[35] Domínguez-García, R., Méndez-Muros, S., Pérez-Curiel, C., & Hinojosa-Becerra, M. (2023). Political polarization and emotion rhetoric in the US presidential transition: A comparative study of Trump and Biden on Twitter and the post-election impact on the public. El Profesional de la información, 32(1).

[36] https://www.bbc.com/news/world-europe-60506682

[37] https://www.nytimes.com/2022/10/27/technology/elon-musk-twitter-deal-complete.html

[38] Yadouleton, A., Sander, A. L., Adewumi, P., de Oliveira Filho, E. F., Tchibozo, C., Hounkanrin, G., René, K. K., Ange, D., Kohoun, R. K., Nari, R. C., Salifou, S., Saizonou, R., Kakai, C. G., Bedié, S. V., Al Onifade, F., Nagel, M., Aïssi, M. A. J., Akogbeto, P., Drosten, C., ... Drexler, J. F. (2022). Emergence of SARS-CoV-2 Delta variant, Benin, May–July 2021. Emerging Infectious Diseases, 28(1), 205-209.

[39] https://www.who.int/publications/m/item/weekly-epidemiological-update---20-october-2020

[40] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9609247/

[41] Altig, D. E., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., ... & Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. Journal of Public Economics, 191, 104274.

[42] Isadora Cerullo (2020). Impact and meaning of the postponement of Tokyo 2020 from an athlete perspective. Emerging Infectious Diseases, 4, 28-36.

[43] Branscome, E. (2021). Colston's Travels, or Should We Talk About Statues? ARENA Journal of Architectural Research, 6(1), 1-16.

[44] Emiliano del Gobbo, Sara Fontanella, A. Sarra, & Lara Fontanella (2020). Emerging Topics in Brexit Debate on Twitter Around the Deadlines. Social Indicators Research, 156, 669-688.

[45] Cranmer, F. (2023). October 2022 to January 2023. Ecclesiastical Law Journal, 25, 247-254.

[46] N. Davies, R. Barnard, C. Jarvis, T. Russell, M. Semple, M. Jit, & W. J. Edmunds (2020). Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. The Lancet. Infectious Diseases, 21, 482-492.

[47] Griffin, S. T. (2020). Covid-19: England will return to regional restrictions amid rapid testing push. BMJ, 371, m4577.

[48] Prosser, C. (2020). The end of the EU affair: the UK general election of 2019. West European Politics, 44(1), 450-461.

[49] Hanlon, H. M., Bernie, D., Carigi, G., & Lowe, J. (2021). Future changes

to high impact weather in the UK. Climatic Change, 166.

[50] R. Thakur & Isha Malhotra (2020). Terrorism as a Media Specific Event: Performative Frames of Uri and Pulwama Reportage in Indian News Media. Media Watch.

[51] Victoria Schofield (2019). Indian General Elections and the State of Jammu and Kashmir. The Round Table, 108(479-480), 479-480.

[52] Roy (2021). Reimagining Citizenship in India Today. PS: Political Science & Politics, 54(2), 631-632.

[53] Š. Ganguly (2021). Is the Deep Freeze Between China and India Turning Hot? Current History, 120(825), 159-161.

[54] J. Anandraj, Y. Krishnamoorthy, P. Sivanantham, Jilisha Gnanadas, S. Kar (2021). Impact of second wave of COVID-19 pandemic on the hesitancy and refusal of COVID-19 vaccination in Puducherry, India: a longitudinal study. Human Vaccines & Immunotherapeutics, 17, 5024-5029.

[55] Suryakant Yadav, P. Yadav, N. Yadav, & C. Yadav (2020). The Peak of COVID-19 in India (Preprint).

[56] Anand Badola (2021). DIGITAL SITES OF PROTEST: FARMERS' PROTEST IN INDIA AND THE CONSTRUCTION OF A COLLECTIVE IDENTITY ON FACEBOOK. AoIR Selected Papers of Internet Research.

[57] https://www.yalemedicine.org/news/covid-timeline