Linköping University | Department of Management and Engineering
Master's thesis, 30 credits| Master's programme
Spring 2024| LIU-IEI-FIL-A--24/04647--SE

Explore weight loss discussions and sentiments on Reddit during the COVID-19 pandemic

Yang Qian

Supervisor: Carl Nordlund
Examiner: Erik Rosenqvist

Linköping University
SE-581 83 Linköping, Sweden
+46 013 28 10 00, www.liu.se

Abstract

This study investigates the weight loss discussions and sentiments on Reddit during the COVID-19 pandemic. This study collected and analyzed a dataset containing 35,384 posts from March 2019 to March 2021 in the *loseit* subreddit. To compare and explore how discussions and sentiments changed within this weight loss subreddit, the analysis utilized two unsupervised machine learning methods, topic modeling and sentiment analysis, to explore the evolution of the topics and emotional expressions of weight loss discussions over time. By employing topic modeling (LDA), the study identified 16 topics within five categories of weight loss discussions: Weight Management, Diet, Physical Exercise, Emotions and Support, and Appearance. However, topic related weight loss as well as upstream and downstream such as Weight change (T2), Calorie tracker (T3), Motivation (T4) and Workout plan (T15) varied over time. Furthermore, Sentiment analysis indicated that discussions in the loseit subreddit are mainly positive rather than negative. There was a slight increase in negative sentiments and a decrease in positive throughout the study period. Emotion analysis indicates the complexity of the emotion dynamic in weight loss discussions. Moreover, users in this weight loss subreddit frequently expressed positive emotions, such as anticipation, and negative emotions, such as fear. However, the sentiments and emotions varied significantly among most topics. These findings reveal how social media affected netizens' attitudes and behaviour on this topic and how it changed during this global crisis. While the study noted significant sentiment changes in some topics across the two time periods in this weight loss subreddit, it remains cautious about concluding the impact of the pandemic in weight loss discussion on social media. Further study could explore a more extended period in multiple subreddits or social media platforms to conclusively link these trends with the pandemic's effect.

Table of Contents

List of tables

List of figures

1.Introduction

2020 was a notable year in the world's recent history. The COVID-19 pandemic has profoundly impacted public health and notable changes in individuals' behavioral patterns worldwide (McKibbin et al., 2020). In response to the pandemic, more than 100 countries implemented stay-at-home orders and social distancing measures. Although these measures controlled the spread of the virus effectively, they also brought various complexities and challenges (Hale et al., 2020; Onyeaka et al., 2021). For example, there have been substantial social and economic losses all over the world (Shah et al., 2020; Lopez-Leon et al., 2021), particularly in sectors such as tourism, entertainment, and hospitality, which resulted in many job losses during the pandemic (Kozicki et al., 2020). In addition to macro-level impacts, these measures have also been reflected in individuals' daily lives, such as physical activity (Tison et al., 2020) and eating habits (Renzo et al., 2020; Bennet et al., 2021). These lifestyle changes further impacted individuals' weight. A few studies have shown that many people reported significant weight gain during the pandemic (Khubchandani et al., 2022; Rosenkilde et al., 2024). At the same time, this weight gain phenomenon has been highly discussed on social media, such as the popular term "quarantine 15" (Keel et al., 2020). This term has become synonymous with netizens making jokes about weight gain while staying at home during the pandemic.

Previous studies have shown that social media platforms were popular tools for sharing health-related information and health interventions, including weight management (Jane et al., 2018). However, most of these studies focus on the outcome of social media on weight management during the pandemic (Schienkiewitz et al., 2022). Not much research has been conducted on how weight loss attitudes and behavior change during a global event such as the COVID-19 pandemic. There are very few studies conducted on this topic. For example, Tang et al. (2022) presented a content analysis of 100 most viewed videos on YouTube about weight loss during quarantine. They found that weight-related graphic posts usually discussed weight loss and resulted in both positive and negative consequence. The data from Google Trends also showed that "weight loss" was one of the most frequently searched topics on the internet during the pandemic (Google Trends, 2024). Furthermore, social media was also an important online space for emotional expression during the public crisis (Ferrara & Yang, 2015). Research has indicated that emotional dynamics on social media could influence netizens' emotional expressions and further shape the public's collective emotional response to a crisis (Papacharissi, 2015). However, in weight management discussions, different netizens could express opposite emotions on social media. Some netizens could express positive emotions such as sharing motivation and successful experiences. Others could express negative emotions and may cause weight bias or stigma on social media (Pearl et al., 2020). This bias could lead to negative self-perception in netizens and even trigger eating disorders such as emotional eating (Mason et al., 2021). Therefore, more research is needed on the discussions and sentiments related to weight loss during the pandemic.

Among these social media platforms, Reddit[1] is one of the largest and most popular, with a unique forum structure of subreddits. Each subreddit (e.g., r/loseit, r/fitness) is an anonymous online community focused

[1] Reddit Inc., https://www.redditinc.com/ (Accessed: Mar 1, 2024)

on a specific topic, such as health and wellness (Britt et al., 2023). Compared to real-time information platforms such as Twitter (X), discussions on Reddit become more comprehensive and in-depth. Moreover, many studies have explored the significant changes in discussions and sentiments on Reddit during the pandemic, especially those related to health behaviors (Wanchoo et al., 2023; Low et al., 2020). This makes Reddit an important site for academic research on health-related studies in the global health crisis.

Based on the above background, the purpose of this study is to conduct exploratory research on how the weight loss discussions and sentiments change during the pandemic, with a specifical focus on a weight loss subreddit on Reddit. By analyzing the contents within this subreddit, I would like to understand the evolution of weight loss behaviors and attitudes on social media in the context of the global health crisis. The following research questions are raised for this study:

> *RQ1: What discussions about weight loss are mainly talked about on Reddit? How did the thematic content of weight loss discussions change from pre-pandemic to during-pandemic era?*
>
> *RQ2: How did sentiment and emotional expressions of weight loss discussions shift across these two periods?*

To address these research questions, this study analyzed the submissions data collected from the *loseit* subreddit. To conduct a comparative study, this dataset is divided into two time periods: before and during the pandemic for analysis. After that, two approaches are utilized to track discussion and sentiment trends over time in the *loseit* subreddit. Specifically, Latent Dirichlet Allocation (LDA) is applied to the dataset to identify key topics and observe how these topics evolve through a temporal analysis. Sentiment analysis is conducted to quantify the overall sentiment polarity and emotion distribution in the subreddit. Lastly, topic sentiment and emotion analysis are used to identify and compare the changes in discussion topics and sentiment across the periods.

By analyzing the discussion and sentiment in this weight loss subreddit on Reddit, this study aims to contribute to several points. First, this study could extend the understanding of the social media dynamic in the context of a global crisis. Second, it could provide new insight into how emotional dynamics affect health behavior during the global health crisis. Third, this research could also offer valuable insights for public policymakers. By understanding users' behavioral and emotional changes on Reddit during the pandemic, they can design and implement more effective health interventions.

This paper is organized as follows: Section 2 reviews the relevant background, including the weight changes during the pandemic, emotional dynamics on social media, Reddit characteristics, and related weight loss research. Section 3 describes the data and methods used in this study, including the data collection, preprocessing, methodological framework, and specific procedures. Section 4 presents the results and interpretation of the findings from the applied approaches. Section 5 is the discussion that delved into the study’s findings. Limitations of this study and future research are also addressed in this section. Finally, Section 6 is the conclusion of the study.

2.Background

2.1 Weight changes during the pandemic

The Covid-19 pandemic had a profound impact on the daily lives of billions of people across the world (WHO, 2020). At the end of 2019, as the pandemic spread globally, most countries issued varying degrees of social lockdown and stay-at-home orders to prevent the spread of the virus. This led to many challenges and changes in lifestyles (Knell et al., 2020) and further affected health behaviours. Many studies have shown significant changes in people's weight during the pandemic. According to a national survey from the American Psychology Association (2021), 61% of adults reported unexpected changes in their weight at the early stage of the pandemic. Moreover, over 42% indicated they gained more weight than expected. Additionally, a questionnaire via Facebook about self-quarantine and weight gain showed that about 22% of adults reported weight gain during the pandemic (Zachery et al., 2020). Previous studies showed that reduced physical activity, changes in eating habits and psychological stress due to lockdown and social isolation were the main factors leading to weight changes during the pandemic (Khubchandani, 2022; AI Zaman, 2023).

Due to the lockdown and social distance measures during the pandemic, many public gym and outdoor exercise facilities were shut down temporarily (Burdett et al., 2021). This led to a significant decrease in people's physical exercise. Serval studies have investigated the impact of the pandemic on the level of physical exercise. A study conducted by Chen and colleagues (2020) showed that despite people trying to exercise at home or choose other indoor activities, overall physical exercise was still reduced significantly due to restrictions of the pandemic. Similar findings were also found in a longitudinal study in Europe. About 40% of participants reported a decrease in the frequency of physical exercise during the pandemic (Cheval et al., 2021). Furthermore, a study by Brand et al. (2020) pointed out that although some inactive participants tried to increase their exercise frequency before the pandemic, 1/3 of the participants still reduced their exercise intensity and workout time during the lockdown. This negative impact on people's physical exercise also correlated with weight changes during the pandemic (Robinson et al., 2021). This lack of long-term exercise during the pandemic may lead to weight gain and other health issues like obesity (Bhutani, 2020).

At the same time, changing people's eating habits during the pandemic (Renzo et al., 2020) also brought challenges to weight management. González-Monroy et al. (2021)'s longitudinal study systematically reviewed how eating habits changed throughout the pandemic, and they found that adherence to healthy diets has decreased. The results showed an increase in the frequency and quantity of food eaten, the consumption of sweets, snacks and processed food, with a decrease in fresh food like fruit and vegetables. These changes in eating behaviour may lead to rapid weight changes, especially weight gain in the short term (Swinburn et al., 2004). In addition, food supply disruptions and supermarket shopping restrictions further worsened the trend towards unhealthy eating habits (Janßen et al., 2021).

The uncertainty of the pandemic and ongoing health threats also affected people's mental health. The mental problems during the pandemic might lead to emotional eating beahvior that causes weight changes. A study shows that due to the home lockdown and limitation of social activity, people turned to high-calorie food to handle the increasing mental pressure and anxiety (Smith et al., 2021). Muscogiuri et al. (2020) also have similar findings that people tend to have high-sugar and high-fat food during the quarantine. Another study in Europe and the U.S. also found that people ate more fast food or snacks for comfort during the lockdown due to anxiety and boredom (Pellegrini et al., 2020). Additionally, a cross-sectional study via an online questionnaire found that during the lockdown, the pressure also affected people's sleep quality and negatively affected their weight (Vanoh et al., 2023).

2.2 Emotion dynamic on social media

As the pandemic spreads, physical interaction between people has been significantly reduced due to social distancing measures and lockdown restrictions. This resulted in an increase in active user activity on social media such as Twitter, Facebook, and Reddit (Chan, Nickson, Rudolph, Lee, & Joynt, 2020; Thelwall & Thelwall, 2020). Social media platforms have quickly became a central online space for information dissemination and personal interaction around the world (Saud et al., 2020). At that time, people were relying on these platforms to obtain information about the pandemic (Cinelli et al., 2020), maintain social connections (Seifert & Hassler, 2020) and seek social support (Nabity-Grover et al., 2020). However, social media was also an important platform for people to express their emotions in public (Naskar et al., 2020), especially in the context of the global events or crisis.

Emotions in social media

Human emotions are very complex to study (Lindquist and Barrett, 2010). Researchers have conducted studies in this area for many years to understand the origins, manifestations, and impacts of emotion on human behavior. Early theories, such as the theory of evolution proposed by Darwin (1872), focused on emotions as adaptive responses necessary for survival. James's theory of emotions (1884) presents that the emotional feelings are the perception of changes occurring in one's body. Over time, emotion theory gradually incorporates psychological, sociological and cultural perspectives. For example, Lazarus's cognitive theory (1991) emphasized the role of cognitive appraisals in emotional experiences. According to this view, emotions arise from the individual's interpretation of events rather than the events themselves. In recent years, with the rise of social media, emotions research has expanded into online space to explore how these emotions are expressed, perceived, and spread in online contexts (Pang & Lee, 2008; Kramer et al., 2014).

According to Plutchik's (1980) *wheel theory of emotions*, emotions are classified into eight primary types: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation with varying intensities, along with two moods: positive and negative. Plutchik proposed that these basic emotions form the foundation of all other complex emotions and are universal across human experiences. As shown in the Figure 1, these emotions are presented in a color wheel. Emotions next to each other share similar characteristics, and those directly across from one another represent opposites. Each primary emotion pairs with an opposite emotion which is joy and sadness, trust and disgust, fear and anger, surprise and anticipation. This theory provide a fundamental structure and framework on emotion study in social media. However, compare to traditional face-to-face communication, the expression of these emotions outlined in Plutchik's (1980) concept is more diverse and instantaneous online. On social media, netizens express their emotional responses in tweets or posts with multiple forms like text, emoticons, images, and videos (Highfield & Leaver, 2016). This emotional expression is usually fluctuates dynamically in response to events such as public health event, natural disasters or sociopolitical movements (Cinelli et al., 2020; Lu et al.,2015; Kimberly et al., 2020).

Previous studies have shown that compared to positive emotions, negative emotions such as fear and sadness are particularly common and widely spread during crisis events. During the pandemic, many researchers have focused on analyzing these emotional dynamics on social media. For example, Lwin et al. (2020) analyzed the trends of emotions with over 20 million tweets worldwide. They found that public emotions during the pandemic shifted rapidly from fear to anger, with notable expressions of sadness and joy. Similarly, Yu et al. (2021) conducted a longitudinal study to examine temporal emotional dynamics on a Chinese social media platform. They found that negative emotions were the most salient emotions detected on Weibo. These findings suggest that social media have become a crucial space where emotions are expressed and amplified through interactions, leading to phenomena such as emotion contagion (Kramer el al., 2014).

Figure 1 - Plutchik's wheel of emotions

Emotions Contagion

The spread of emotions is significant in the context of public health crises. Emotion Contagion Theory (Barsade, 2002) provides a framework for understanding how emotions spread among individuals through social networks. This theory is originated from psychology and as part of the broader concept of social contagion. At the beginning, ECT theory was used to explain emotion exchanges in traditional face-to-face interactions (Hatfield et al., 1993). It states that when individuals interact with others, they tend to intimate others' emotional expressions, such as facial expressions, speech, and body posture or movements, until their emotions align. However, with the rise of social media, the application of this theory has extended to online platforms (Krone et al., 2018). For example, an individual expressing emotion on social media can quickly influence their network and shape the emotional tone of entire groups (Bringmann et al., 2018; Fan et al., 2018). This process of emotion contagion involves more complex interactions than initially described by the theory. Unlike traditional face-to-face interaction, emotion cognition on social media is not limited by time or space. Anyone can comment, like or retweet, which could intensify and spread emotion rapidly (Papacharissi, 2016). The literature suggests that the role of emotional contagion during a public crisis is crucial (Steinert, 2021). Emotional dynamics on social media influence users' individual emotional experiences and shape the public's collective response to a crisis on a larger scale (Smith et al., 2014). These collective emotions can deepen the cognitive processing and reflection of social events, sometimes they could even change the public perceptions and emotional responses (Perelló-Sobrepere, 2017).

2.3 Reddit and related work

Studies have shown that online interaction has a strong relationship with the health outcome. Social media has been used to research on different health conditions, such as mental health (Primack et al., 2017) and chronic diseases (Patel et al., 2015). Compared to other social media, such as Twitter, Facebook or other popular platforms, Reddit is more suitable for health research due to its unique forum structure and anonymity (van der Nagel, 2013). Since found in 2005, Reddit has grown into one of the largest social media platforms in the world. Its unique subreddit structure is particularly suitable for discussing around specific topics by neitenzes (van der Nagel, 2013). In these subreddits, netizens can share their experiences, ask questions and get feedback. Moreover, Reddit's design focus on community self-discipline and this interaction form is less common on other social media platforms such as Twitter or Facebook (Massanari, 2015). This anonymous characteristic encourages netizens to be more open to discuss sensitive topics such as substance use disorder (Chi et al., 2023). However, this anonymity also brings challenges, such as toxic language, to exist on Reddit as much research has been done in this area (Mohan et al., 2017; Xia et al., 2020; Almerekhi et al., 2020). Additionally, Reddit's voting system (upvote/downvote) allows netizens to judge the posts and comments. This system makes the most valuable content easier to discover and promote (Gaudette et al., 2018).

Various studies show that the pandemic outbreak significantly affected health-related discussions and sentiment on Reddit on these changes. Wanchoo et al. (2023) used NLP on 1 million posts in 22 subreddits to compare language marker changes related to diet, physical activity, substance use, and smoking. They found that the focus of discussions shifted from daily life to topics related to health and emotional management during the pandemic. Moreover, Reddit is also widely used to study mental health. For example, Low et al. (2020) also used natural language processing and machine learning to analyze comments from mental health communities on Reddit, such as r/anxiety, r/depression, and r/SuicideWatch. They found that isolation had the most significant impact on mental health during the pandemic. In eating behavior, several studies have looked specifically at discussions of eating disorders and emotional eating. For example, Amin & Liu (2024) explore the eating disorder on Reddit through a comparative analysis of the subreddits r/AnorexiaNervosa, r/BingeEatingDisorder, and r/EatingDisorders and found that the pandemic has resulted in significant changes in discussion topics and online support-seeking behaviors related to eating disorders.

However, although there have been many health studies, such as physical activity and dietary patterns on Reddit, few specific studies on specific analyses of weight management. In particular, changes in discussions and sentiments about weight management behaviours were still unknown during the pandemic. Many previous studies have focused on the association between pandemics and weight change. Several studies have explored the impact of the pandemic on weight management through online questionnaires or electronic surveys (Almandoz et al., 2022; Barenbaum et al., 2022). These studies have consistently found that pandemics have a negative impact on these populations, often leading to weight gain (Zachary et al., 2020; Glazer & Vallis, 2022). For example, a study by Caldwell et al. (2022) found that people with overweight and obesity experienced increased stress during the pandemic, which severely impacted their weight management behaviours and led to more pronounced mental health problems. Kuk et al. (2021) came to similar conclusions, reporting that approximately half of the participants noted overall health status during the pandemic deterioration, including mental health, physical activity, and eating habits. Additionally, research has focused on the impact of social factors during the pandemic, such as weight bias and stigma, which is particularly heightened on social media (Pearl et al., 2020; Flint et al., 2020). This relates to some posts about quarantine-15. Based on the above background, although these studies have been conducted to provide valuable insights into the challenges of weight management during the pandemic, there is still a lack of in-depth analyses of weight loss discussions on social media platforms. This study aims to fill this gap by analysing posts on Reddit through the NLP approach to explore the changes in online discussions and sentiments about weight loss during the global crisis.

3.Data and Method

This section outlines the study's data and methodology, beginning with section 3.1, which describes the Reddit dataset and the collection process. Section 3.2 covers the preprocessing of the Reddit dataset. Section 3.3 provides an overview and details of the analytical techniques employed in the study. Sections 3.4 and 3.5 discuss the details of the topic modeling (LDA) and sentiment analysis (Vader and NRC-EIL) techniques used. Ethical considerations related to data collection are addressed in section 3.6.

3.1 Data collection

This study drew the threads from the *loseit* subreddit on Reddit as the primary data resource. This subreddit was chosen as the data source for this study for multiple reasons. Firstly, *loseit* is the largest weight-loss subreddit on Reddit, with over 4 million members (Reddit, 2024). Second, given the *loseit* subreddit provides support and discussion about weight loss, users in this subreddit might have the weight-related health impact of the pandemic. Additionally, previous studies have been conducted to analyze the research weight management in this subreddit (Pappa et al.,2017; Liu & Yin, 2020; Hwang et al., 2020). Thus, this study analyzed the *loseit* dataset to capture the discourse and sentiments discussed and expressed about weight loss behavior.

In this study, the *loseit* subreddit data for analysis contains submissions' texts and related metadata. I initially planned to collect the *loseit* data using the Reddit API during the collection process. However, due to the limitation of Reddit API, it only allowed access to submissions which is categorized as "Hot", "New", "Top" or "Rising". This method did not support retrieving past data directly for specific periods, which was crucial for my research covering periods before and during the pandemic. To solve this, I turned to a historical dataset of Reddit. The historical dataset is publicly available in JSON format through Pushshift archives via Academic Torrents[2], covering the duration from June 2005 to Dec 2023. After inspecting the dataset, it was found that some missing data, such as deleted or removed submissions, were also included. To ensure completeness and accuracy of the further analysis, it was necessary to exclude those submissions and comments that may have been deleted or removed during the study period. Therefore, I extracted the submission IDs of the relevant *loseit* submissions from this historical dataset and then used API to scrape these submissions. With these IDs, collecting the relevant submissions' selftexts and related metadata through the API was feasible. 115,123 submission IDs related to the *loseit* subreddit were extracted. This data-collecting approach focuses on scraping the additional metadata not present in the historical dataset and ensuring any deleted or removed submissions after the dataset's latest update was accounted for.

115,123 submission IDs were then used to scrape the up-to-date submissions via the Reddit API. The *loseit* data for this study covered the period from March 2019 to March 2021, encompassing the pandemic's key outbreak date. This timeframe was simplified into two periods to compare the subreddit's activity across these periods. Specially, the period from March 01, 2019 to March 10, 2020 is defined as the pre-pandemic

[2] Academic Torrent, https://academictorrents.com (Accessed: Feb 10, 2024)

era, while the period from March 11, 2020 to March 31, 2021 is defined as the era during the pandemic. Although some countries started lockdown earlier than March 11, 2020 (Koh, 2020), I chose to use this time division based on the World Health Organization (WHO) announcement. On March 11, 2020, WHO declared COVID-19 a global pandemic (WHO, 2020). This division might not fully capture the timelines of some specific countries. However, given that the purpose of the study was to analyze global trends in discussion and sentiment about weight loss, this division provided a benchmark for the analysis. Then, a data scraping process with the PRAW (Python Reddit API Wrapper) package was designed and applied to scrape submissions from the *loseit* subreddit. Given the large dataset volume and the Reddit API limit of 100 queries per minute, a Python script was written to collect Reddit data efficiently while complying with the "Reddit API Terms of Use" (Reddit, 2024). Once all scraping is completed, these CSV files containing all the 115,123 submissions are combined into a single CSV file automatically. Figure 2 illustrates the data collection process used in this study.

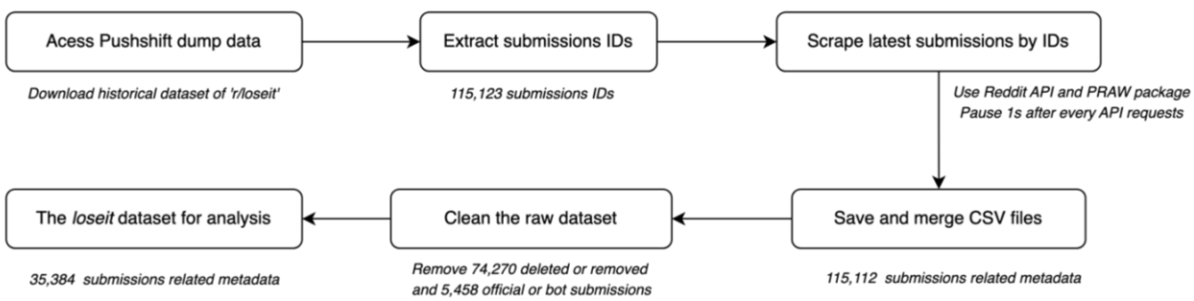

Figure 2 – The flowchart of Reddit data collection process

However, 11 submission IDs could not be matched with the relevant post during the collecting process. As a result, only 115,112 submissions were collected successfully. After that, a data cleaning process was employed to remove 74,270 deleted or removed submissions. In addition, 5,458 posts from official bot accounts like 'AutoModerator' and official weight loss challenges such as 'European Accountability Challenge' from this subreddit were also removed in this step. So the process retained 35,384 submissions, with 16,244 (46%) from before the pandemic and 19,140 (54%) from during the pandemic. This *loseit* dataset contains the metadata including "Creation time", "Title", "Username", "Selftext", "Score", "Number of comments", "URL" and "Period". Notably, "Period" was divided based on the WHO timeline into two distinct phases named 'pre-pandemic' and 'dur-pandemic' for further analysis. Table 1 below presents two popular submissions with higher scores and comments in the *loseit* dataset for each period:

Creation Time	*Title*	*Author*	*Selftext*	*Score*	*Number of comments*	*URL*	*Period*
2019-03-12 15:27:24	"New ID required. Reason: Weight loss".…	username	For the past year or so, every time.…	21868	473	https://www. reddit.com/ r/loseit/...	pre_pandemic
2020-03-22 00:12:37	I'm an RN taking care of COVID-19 patients…	username	This is a throwaway account to…	29405	788	https://www. reddit.com/ r/loseit/...	dur_pandemic

Table 1 - Example of submissions and related metadata per period in the loseit dataset

- **Creation Time:** The specific time and date of each post created.
- **Title:** The title of each post.
- **Author:** The username on Reddit who posted the post.
- **Selftext:** The content of each post.
- **Score:** The differences between the likes and dislikes. This reflects the quality of each post.
- **Number of comments:** The number of comments under a post. This indicates the popularity of each post.
- **URL:** The link to the original post.

	Pre-pandemic			**Dur-pandemic**		
	Mean	*Median*	*SD*	*Mean*	*Median*	*SD*
Posts per day	43.2	42.0	10.8	49.6	49.0	11.6
Scores per post	128.4	7.0	620.8	134.7	7.0	698.7
Words per post	281.7	216.0	230.2	275.4	212.0	216.8

Table 2 - Basic statistics overview of the loseit dataset before and during the pandemic era.

Table 2 provides an overview of the basic statistics for the *loseit* dataset in two periods. This includes the mean, median, standard deviation (SD), the number of daily posts, and the length of each post in the *loseit* subreddit. From this table, the daily number of posts has increased during the pandemic. This indicates that a slight increase in subreddit activity throughout the study period. Furthermore, regarding the scores per post, although the median score has retained at 7.0, we could see that the average score of each post has increased during the pandemic. This indicates that some posts during the pandemic received higher scores than those in the pre-pandemic era. Additionally, the average word count per post has decreased slightly during the pandemic, but the change was not substantial.

Additionally, I also tracked and visualized monthly trends in the submissions posted. As Figure 3 shows, there was a significant increase in the number of posts in the *loseit* subreddit starting from March 2020, about 2 months after WHO officially declared the pandemic. This increase continued until August 2020, after that, it returned to more typical levels.

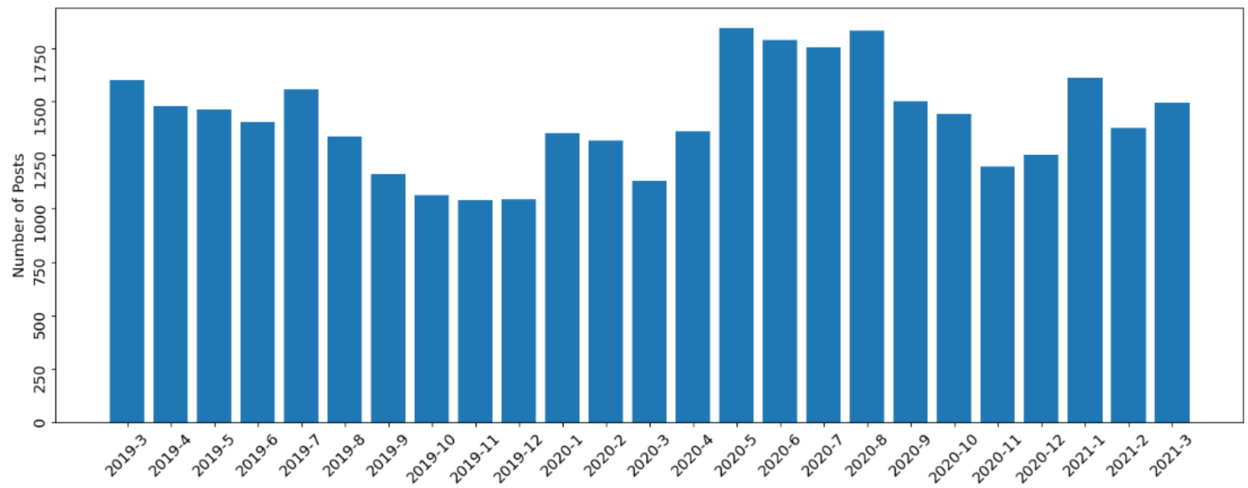

Figure 3 - Number of the posts posted per month from March 2019 to March 2021

3.2 Data processing

After the data collection, the next step is to preprocess the text from submissions and comments for further analysis. The raw text data from social media are often filled with irrelevant information and noise (Miao et al., 2019). Thus, this preprocessing aims to clean up the data to improve the accuracy and efficiency of future machine learning model applications. The cleaning step involved the following steps: 1) removing HTML tags, 2) removing punctuation marks and some special characters such as line breaks "\n" and the Zero Width Space character "&#x200B", 3) converting to lowercase and 4) removing numbers and other non-alphabetic characters. Moreover, although most of the text was in English, some non-English texts might be presented in the *loseit* dataset. To address this, I also removed non-English text to keep focus on English content only. After that, the cleaned text data was prepared to be preprocessed. The preprocessing steps included 1) removing common stop words, 2) normalizing the text, and 3) tokenizing the text. Some common English stop words like "the", "is", and "at" were removed in this step. These words generally could not provide helpful information about the themes and emotions of the text. Regarding text normalization, this study chose to lemmatize the text rather than stemming the text for more accurate analysis. Two popular natural language processing (NLP) libraries in Python, NLTK (Natural Language Toolkit) and spaCy, were utilized to normalize the text. NLTK is one of the oldest and most comprehensive Python libraries for natural language processing (Bird et al., 2009), while spaCy is a newer and rapidly growing library designed for production use (Honnibal et al., 2020). These two techniques were applied to reduce the effect of part-of-speech variations on the NLP models and ensure the preservation of the original meaning of the words. This process involves converting words to their base or original forms. For example, in this *loseit* dataset, words such as "felt", "struggling", and "calories" would be normalized to "feel," "struggle," and "calorie," respectively. Additionally, I only kept nouns, verbs, adjectives, pronouns and adverbs to reduce the word sparsity for topic modeling. Words with fewer than three characters were excluded during this step since they contributed little to the model. However, this preprocessing step only removes HTML tags and special characters for the VADER analysis. Elements like punctuation marks (e.g.!!!) and capital letters (e.g., "I AM GOOD,") are considered strong sentiment intensity in the VADER analysis.

3.3 Analytical strategy

This research design consists of five steps: data collection, preprocessing, modeling, analysis, and result interpretation. The details of the first two steps were already discussed in the previous sections, 3.1 and 3.2. In the modeling step, since this study mainly focuses on exploring how discussions and emotions evolved in the *loseit* subreddit from pre-pandemic to dur-pandemic eras, I chose to utilize topic modeling and sentiment analysis methods. Specifically, Latent Dirichlet Allocation (LDA) and two lexicon-based techniques (VADER & NRC-EIL) were employed. Figure 4 shows the entire research design flowchart of the study.

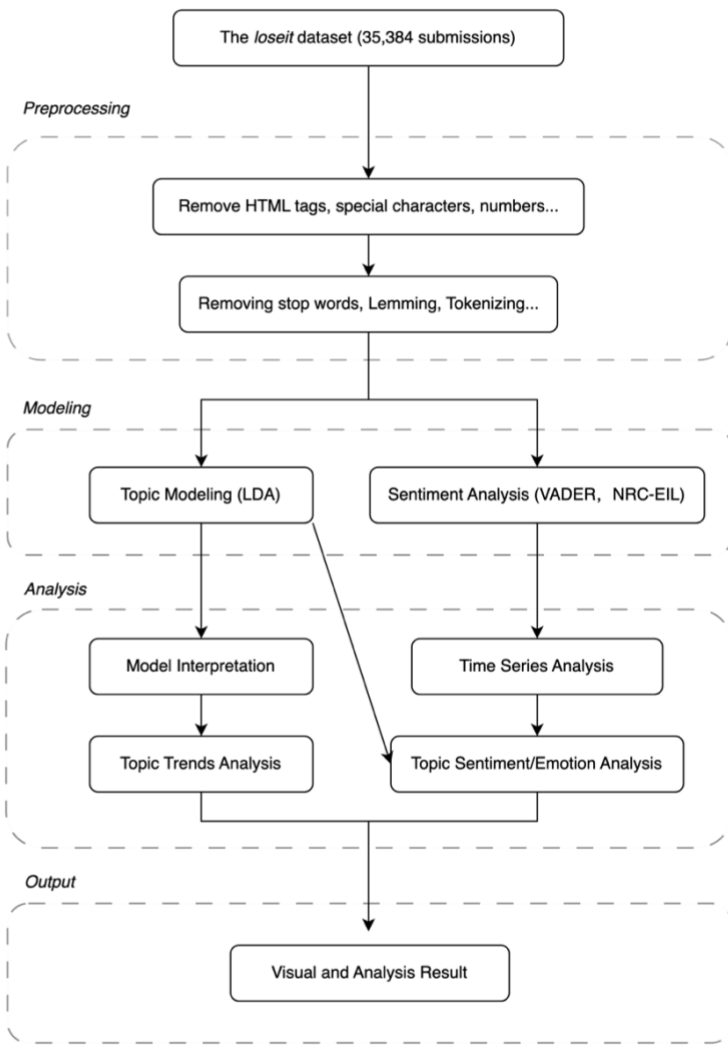

Figure 4 - The research desigh diagram

In the topic model path, an LDA model was applied to the *loseit* submissions. The composition of topics could be identified in the *loseit* subreddit over a specified period. After that, I used a qualitative method to interpret the LDA topic result of this subreddit. This step included interpreting the top 10 keywords of each topic generated from the LDA algorithm in this subreddit. And then, I used a topic trend analysis by calculating the monthly distribution of LDA topics. By observing the proportion of posts for each topic over time, I could identify the changes in the focus of discussions on specific topics. Conversely, in the sentiment analysis path, I conducted VADER & NRC-EIL to classify and categorize the sentiment and emotions expressed in the *loseit* subreddit. VADER (Valence Aware Dictionary and Sentiment Reasoning) was used to obtain basic sentiment scores for each post, and NRC-EIL (NRC Emotion Intensity Lexicon) was used for a more detailed emotion analysis. Furthermore, a time series analysis provided a monthly breakdown of sentiment and emotion trends. After that, I analyzed each topic's sentiment polarity and emotion distribution for both pre-pandemic and during-pandemic periods. To deepen this analysis, I compared the proportions of different sentiments (positive, negative, or neutral) within each latent topic across these two periods. A statistical test was performed to determine if there were significant differences in each topic's sentiment across the periods. Details about modeling and analysis are discussed in the following sections 3.4 and 3.5.

3.4 Topic modeling

3.4.1 Latent Dirichlet Allocation

This study utilized topic modeling to explore the hidden structure within the discussion of the *loseit* subreddit during the study period. While Dynamic Topic Modeling (DTM) can provide more insights and details into the evolution of topics over time (Blei et al., 2006), LDA was chosen for this study instead. The reason for this choice was that the research aimed to capture the overall thematic trends and distributions within the subreddit rather than tracking the internal changes of each topic. Thus, the LDA technique is more appropriate for the research objectives in this study. LDA is a statistical model that helps discover various topics and keywords within a large text volume (Blei et al., 2003). It is the most popular topic modeling technique, and its strength is in identifying latent topics within collections of text. LDA can discover latent topic distributions by examining keyword co-occurrence within documents. As an unsupervised machine learning technique, it assumes that documents are mixtures of multiple topics and that topics are mixtures of various words. Figure 5 shows the probabilistic graphical model representation of the Latent Dirichlet Allocation (LDA) model based on Blei et al. paper.

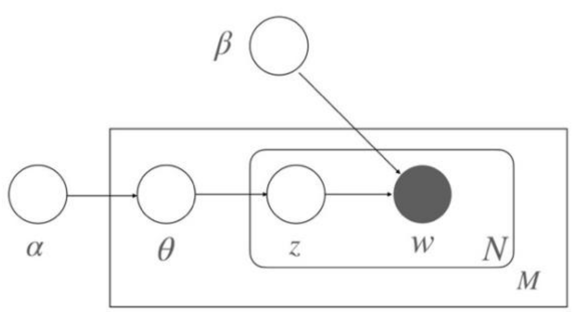

Figure 5 - Graphical model representation of LDA by Blei et al.

As the figure illustrates the LDA generative process:

1. Choose the topic distribution θ_d for document d from Dir (α) $\theta_d \sim \text{Dir}(\alpha)$
2. Choose the topic-specifcic word distribution ϕ_k for for topic k from Dir (β): $\phi_k \sim \text{Dir}(\beta)$
3. For each word in document d:

    3a. Choose a topic $z_{d,n}$ from Multinomial (θ_d): $z_{d,n} \sim$ Multinomial(θ_d)

    3b. Generate a word $w_{d,n}$ from Multinomial ($\phi_{z_{d,n}}$): $w_{d,n} \sim$ Multinomial($\phi_{z_{d,n}}$)

θ_d represents the topic distribution for document d, sampled from a Dirichlet distribution with parameter α. ϕ_k represents the topic-specific word distribution for topic k, where k is the topic assigned to the word $w_{d,n}$. $z_{d,n}$ represents the topic assigned to the n-th word in document d, sampled from the multinomial distribution defined by the topic distribution θ_d. $\phi_{z_{d,n}}$ is the word distribution for topic $z_{d,n}$, which is sampled from a Dirichlet distribution with parameter β.

Figure 6 below presents the article "Seeking Life's Bare (Genetic) Necessities" by Blei and his colleagues for LDA. This article is a good example of how these topics could be identified through the distribution of words in the LDA model. In this example, we could see that some words frequently appearing are highlighted in different colours. However, words with the same colour represent a distinct topic. For instance, keywords such as "genes", "dna", and "genetic" in yellow colour might be mainly about *genetics*. In contrast, keywords such as "data", "number", and "computer" in blue colour might be related to *data analysis*.

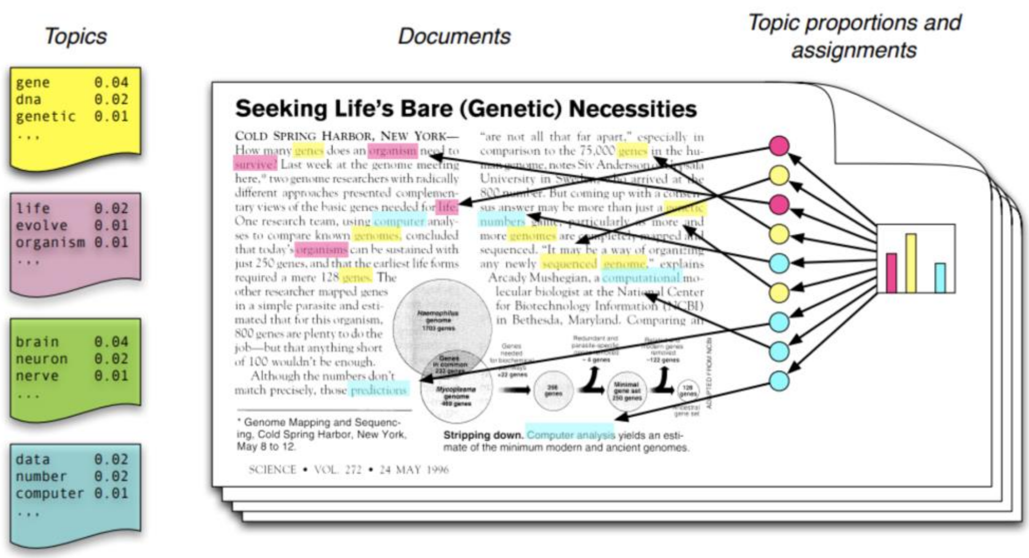

Figure 6 - Example of LDA Applied to the Article 'Seeking Life's Bare (Genetic) Necessities' by Blei et al.

3.4.2 Optimal number of topics

An important aspect of configuring the LDA model is to select a suitable number of topics k (Stevens et al., 2012). Consistency scores are commonly used in evaluating the quality and explainability of topic model algorithms like LDA (Röder et al., 2015). Grimmer & Stewart (2017) have conducted a study that the coherence score could distinguish between topics that are easily interpretable by humans and those that are only meaningful from a statistical standpoint. Models with higher consistency scores indicate that the topics are well differentiated and cohesive. Given a topic k with its top M frequent words denoted as $W = \{ w_1, w_2 , \dots , w_M\}$, the consistency score is typically calculated based on the pairwise coherence of these words. The calculation of the coherence score is the average of the Normalized Pointwise Mutual Information (NPMI) of all words. Below is the equation of the coherence score:

$$Coherence(W) = \frac{1}{(\frac{M}{2})} \sum_{1 \le i < j \le M} NPMI(w_i, w_j)$$

Here, W is the set of the top M words in a given topic. $(\frac{M}{2})$ is the number of unique word pairs that can be formed from M words. This is given by $(\frac{M(M-1)}{2})$. $NPMI(w_i, w_j)$ is the NPMI score for each pair of words (w_i, w_j) in the topic.

To determine the optimal number of topics, I trained LDA models with different topic numbers ranging from 5 to 30 and then examined the coherence score of each model. As shown in Figure 7, the coherence score of topic numbers increased with the number of topics. It peaked at 27 topics with the highest coherence score of 0.408 and started to decline after that. However, I selected the topic numbers from 15 to 20 as the potential optimal topic numbers for the final LDA model. Although these models had lower scores than those with 27 topics, they could provide a more manageable number of topics for in-depth examination and interpretation.

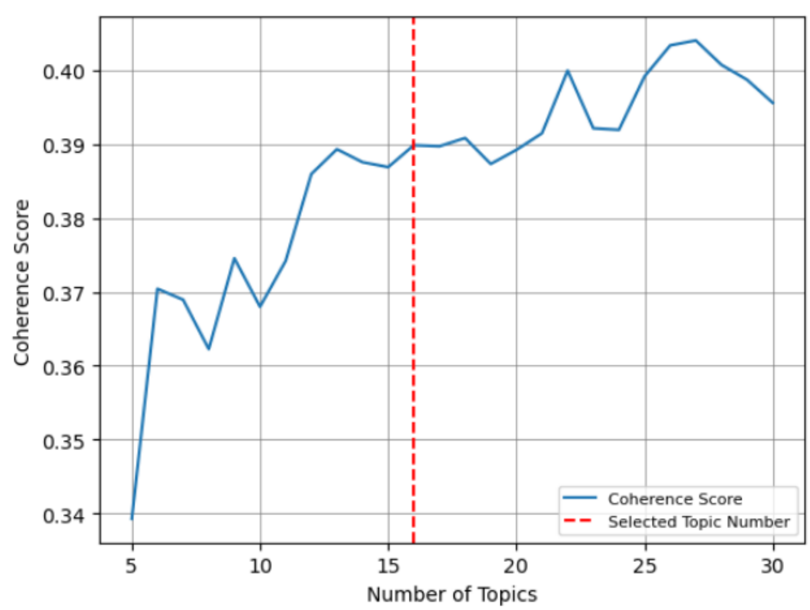

Figure 7 - Topic coherence scores across different numbers of topic ranging from 5 to 30

Table 3 presents the coherence scores corresponding to topic numbers 15 to 20. As shown in Table 3, there is little variation in coherence scores between 15 and 20. This means that there are slight differences in model quality with these topic numbers. Therefore, I finally chose 16 topics for the final LDA model in this study.

Topic number	15	16	17	18	19	20
Coherence score	0.387	0.390	0.390	0.391	0.387	0.389

Table 3 - Coherence scores coresponding to topic numbers 15 to 20

When running the LDA model with 16 topics, I also found that some certain words contributed little to the model. Therefore, I added more custom stop words such as "get", "make", "take", "really", "still", and "even" during this step. After finishing the model tuning, each topic is interpreted and labeled manually based on the most frequent words within each topic. For instance, a topic characterized by frequent keywords of "weight", "lose", and "pound" could be labeled as a topic named "weight change", which focuses on the discussion of personal weight loss progress in the subreddit.

3.4.3 Topic trend analysis

After generating the LDA results, I applied a time series analysis to examine the trends of each topic throughout the study period. Since each post (document) in the *loseit* dataset included timestamps, it was feasible to aggregate these posts by month. I calculated the number of posts and the proportions of each topic per month. This proportion could represent the intensity of the discussion on that topic during that month. By obtaining the monthly topic proportions, I could analyze and visualize the trends of each topic over 24 months to identify which topic changes significantly in the *loseit* subreddit.

3.5 Sentiment analysis

To explore the sentiment changes in the *loseit* subreddit before and during the pandemic, I conducted two lexicon-based sentiment analyzers for sentiment analysis. Basically, the lexicon-based technique uses predefined dictionaries containing words and related sentiment scores to assess the emotional tone of texts (Nandwani et al., 2021). While the lexicon-based technique has some limitations, such as restricted vocabulary coverage and difficulty in capturing sarcasm or puns in sentiment, it is widely applicable to social media posts or movie reviews (Jurek et al., 2015). Unlike machine learning approaches, lexicon-based techniques are straightforward to implement. It doesn't require annotated data in advance and involves any complex model training and parameter tuning as well. For example, a sentiment dictionary such as VADER is designed for social media texts due to its sensitivity to both the polarity (positive/negative) and the intensity (strength) of emotions (Hutto et al., 2014), which is particularly suitable for analyzing Reddit posts in this study.

3.5.1 VADER analysis

In this study, the VADER algorithm is used to calculate the basic sentiment scores for each post. VADER, proposed by Hutto and Gilbert (2014), is a rule-based sentiment analysis tool specifically designed to analyze social media texts. Each word in the VADER lexicon is assigned a predefined valence score, reflecting its degree of positivity or negativity. The valence score ranges from -4 to +4, where -4 indicates the most negative sentiment and +4 indicates the most positive sentiment. Table 4 shows example words and their corresponding scores from the VADER lexicon. As shown in Table 4, words such as "happy", "excellent", and "love" with high sentiment are positive, while words like "sad", "terrible", and "hate" with negative sentiment are negative. Words like "book" and "chair" are neutral.

VADER Lexicon								
Word	happy	excellent	love	sad	terrible	hate	book	chair
Score	2.8	3.4	3.2	-2.5	-3.1	-3.2	0.0	0.0

Table 4 - Example of Vader lexicons

In the empirical analysis, sentiment scores are calculated using the 'polarity_scores()' function from the VADER module. This function determines the text's sentiment polarity, positive, negative, or neutral, and these corresponding sentiment scores. These scores range from 0 to 1 and indicate the proportion of each sentiment in the text. For example, if a post outputs a positive sentiment score of 0.3, 30% of this post expresses positive. Additionally, VADER also calculates a compound score. This score ranges from -1 to +1 and represents the overall sentiment of a text. In this study, after testing various compound score thresholds of 0.05, 0.1, 0.15, and 0.2, I adjusted to 0.15 to achieve a more balanced sentiment distribution. Thus, if the compound score is above 0.15, the text is considered to have an overall positive sentiment.

Figure 8 presents an example of posts in the *loseit* dataset with its corresponding sentiment polarity and scores determined by VADER lexicon. We could see that this is a positive post with a compound score of 0.9633.

```
I'm proud because in the past this was always the start of a pattern: If I missed gym once I missed it a second time because the perfect we
ek was ruined anyways. Then I often would not go back to the gym at all. But not this time. I just called my gym buddy and told him to meet
two days later at the gym so we could work out together. I found it really easy to stay consistent when there is someone who expects me to
show up. But there is more success! My gym buddy who I relied on in the first weeks worked day shifts for 2 weeks straight now, so he could
not work out when I did. But I still got my workout done, alone, with my own willpower! I also sticked to my meal plan and lost some pounds
because the success in the gym motivated me.

Sentiment: positive, Positive Score: 0.16, Negative Score: 0.064, Neutral Score: 0.776, Compound Score: 0.9633
```

Figure 8 - An example of post and corrseponding sentiment output by Vader lexicons

3.5.2 NRC-EIL analysis

The NRC-EIL algorithm is utilized to conduct an in-depth examination of sentiment changes within posts in the *loseit* subreddit. This lexicon consists of nearly 10,000 entries covering eight basic emotions identified by Plutchik (1980). This lexicon assigns specific words in the posts to various emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) with corresponding intensity scores. While *joy*, *trust*, *anticipation*, and *surprise* can be positive or neutral, *fear*, *sadness*, *disgust*, and *anger* are typically negative (Mohammad et al., 2018). Table 5 illustrates example words and their corresponding emotions from the NRC-EIL lexicon. As shown in Table 5, each word from the NRC-EIL is associated with a specific emotion. For example, a word such as "happy" is defined as *joy* while the word "gloomy" is defined as *sadness*. Similarly, a word such as "confident" is defined as *trust*, while the word "afraid" is defined as *fear*.

NRC-EIL Lexicon								
Emotion	Joy	Trust	Anticipation	Surprise	Fear	Sadness	Disgust	Anger
Word	happy	confident	hopeful	shocked	afraid	gloomy	sickened	furious

Table 5 - Example of NRC-EIL lexicons

Figure 9 provides the same example (Figure 6) but with corresponding emotion intensity determined by the NRC-EIL lexicon. We could see that this post contains more positive emotions like *trust*, *joy*, and *anticipation* than negative emotions like *fear*, *sadness*, *anger*, *and disgust*.

```
I'm proud because in the past this was always the start of a pattern: If I missed gym once I missed it a second time because the perfect we
ek was ruined anyways. Then I often would not go back to the gym at all. But not this time. I just called my gym buddy and told him to meet
two days later at the gym so we could work out together. I found it really easy to stay consistent when there is someone who expects me to
show up. But there is more success! My gym buddy who I relied on in the first weeks worked day shifts for 2 weeks straight now, so he could
not work out when I did. But I still got my workout done, alone, with my own willpower! I also sticked to my meal plan and lost some pounds
because the success in the gym motivated me.

Trust : 6, Fear : 1, Sadness : 2, Anger : 1, Surprise : 0, Disgust : 1, Joy : 6, Anticipation : 8
```

Figure 9 - An example of post and corrseponding emotion output by NRC-EIL lexicons

After obtaining the VADER sentiment scores and NRC-EIL emotion lexicon results for each post, I conducted the time series analysis to examine monthly sentiment polarity trends and emotion distribution trends throughout the study period. To do this, I calculated the monthly number of posts categorized as positive, negative, and neutral and the monthly proportion of words associated with each of the eight emotions to track sentiment and emotion trends over time, respectively.

3.5.3 Topic sentiment and emotion analysis

After obtaining each post's sentiment and emotion results, I could associate the LDA topics with them for further analysis which inspired by Monova et al.'s study (2024). By correlating topics with sentiments and emotions, these tendencies associated with different topics could be compared in the *loseit* subreddit over time For example, a topic related to diet might contain more positive sentiments and emotions, such as *joy* and *trust*. *In contrast*, another topic related to appearance might contain more negative sentiments and emotions, such as *fear* and *sadness*. For the process, the LDA model extracts topics from posts to determine popular topics. Then, VADER and NRC-EIL are applied to analyze the frequency and intensity of emotions associated with words in each topic. Finally, a statistical test is used to identify whether there are significant sentiment changes in these topics during the pandemic.

3.5.4 Two proportion Z-test

The two-proportion z-test is used to analyze sentiment changes within the *loseit* subreddit before and during the pandemic. This statistical test helps determine whether the pandemic has significantly impacted the sentiments by comparing the proportions of different sentiment categories (e.g., positive, negative) between the two periods. In this study, we considered the sentiment proportion data collected before and during the pandemic as two independent samples. I conducted two proportion z-test on each positive and negative proportion data. The test involves calculating the proportion of posts within each sentiment category for both periods and then assessing whether the difference in these proportions is statistically significant. The two proportion z-test is calculated using the following formula:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Here, p is the total proportion of the sentiment across both periods. p_1 and p_2 are the proportions of posts with a specific sentiment (positive or negative) before and during the pandemic. n_1 and n_2 are the total number of posts in each period.

The null hypothesis H_0 $(p_1 = p_2)$ assumes no difference in the proportions of sentiments between the two periods. The null hypothesis is rejected if the p-value obtained from the test is less than the predetermined significance level (~0.05). This indicates that there has been a significant change in the sentiment expressions in this weight loss subreddit across the two periods.

3.6 Ethical considerations

Data collection and analysis were conducted in compliance with Reddit's terms of service and the developer's guidelines (Reddit, 2024). All usernames and other potential personal information were excluded before the analysis step. Furthermore, this study only used publicly available historical datasets from Reddit, which are openly accessible and have been used in prior academic research (Chen et al., 2024; Dhankar et al., 2023; Alipour et al., 2024).

4.Result

This section outlines the study's results, presenting the outcomes of the LDA topic model and lexicon-based sentiment analysis. Section 4.1 includes the description and interpretation of 16 LDA topics and the temporal evolution of these topics over time. Section 4.2 includes Vader and NRC-EIL as well as a topic analysis throughout the study period. Finally, a statistical test is utilized to identify the significant sentiment difference for each period.

4.1 Results of Topic modeling

4.1.1 Description of LDA topics

16 topics were generated from the LDA model in the *loseit* subreddit. These topics have been manually named by the researcher into *Clothe fit* (T0), *Daily updates* (T1), *Weight change* (T2), *Calorie tracker* (T3), *Motivation* (T4), *Fitness App* (T5), *Exercise routine* (T6), *Body image* (T7), *Seeking advice* (T8), *Medication* (T9), *Emotional eating* (T10), *Weight loss goal* (T11), *Food choice* (T12), *Negative feeelings* (T13), *Diet control* (T14) and *Workout plan* (T15). Some of these topic names are based on the research in the same subreddit (Liu et al., 2020; Yang et al., 2023). As shown in Figure 10, the number of documents in each LDA topic is sorted from highest to lowest.

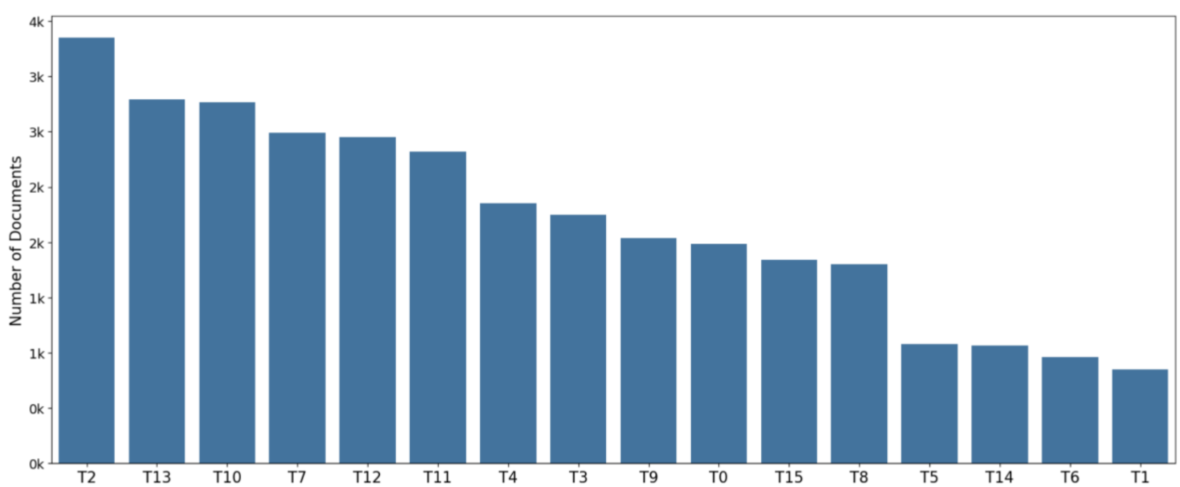

Figure 10 - Number of documents (posts) in each topics (LDA)

Based on the result of the LDA model illustrated in Table 6, we could see that in the *loseit* subreddit, each topic reflects different aspects of the discussion on weight loss methods and healthy lifestyles posted during the study period. These topics have distinct focuses in areas such as weight management, diet, physical exercise and more. Therefore, I categorised these topics into five groups corresponding to specific areas and then interpreted each based on these categories. I also included some examples of posts for each topic to better understand the discussion within each topic. In addition, since most of the posts on Reddit are lengthy, I would select specific paragraphs from posts that better showcase or represent the topic.

Topics		***Keywords***	***Size***
1. Weight Management			**27.1%**
T2	Weight change	weight, lose, lbs, pound, week, gain, month, start, day, loss	10.9%
T11	Weight loss goal	goal, want, back, time, keep, journey, year, post, weight, lbs	8.0%
T9	Medication	year, weight, start, back, gain, high, doctor, lbs, lose, school	5.8%
T1	Daily updates	day, week, today, morning, weight, weekend, stay, time, month, calorie	2.4%
2. Diet			**26.8%**
T10	Emotional eating	eat, food, feel, binge, day, healthy, think, hungry, calorie, want	9.2%
T12	Food choice	eat, meal, food, dinner, snack, lunch, cook, calorie, breakfast, chicken	8.3%
T3	Calorie tracker	calorie, day, eat, weight, deficit, week, fat, burn, scale, muscle	6.3%
T14	Diet control	drink, water, day, cut, sugar, eat, diet, lot, fast, soda	3.0%
3. Physical Exercise			**17.6%**
T4	Motivation	feel, work, back, time, start, day, know, well, hard, month	6.6%
T15	Workout plan	gym, workout, exercise, start, time, work, cardio, routine, year, home	5.2%
T5	Fitness App	calorie, use, day, app, protein, track, daily, body, fat, log	3.1%
T6	Exercise routine	run, walk, minute, mile, week, day, time, leg, start, long	2.7%
4. Emotions and Support			**14.4%**
T13	Negative feelings	life, people, health, time, say, know, family, want, friend, mental	9.3%
T8	Seeking advice	help, diet, thank, advice, know, weight, want, need, start, lose	5.1%
5. Appearance			**14.1%**
T7	Body image	look, body, weight, fat, lose, feel, want, know, always, think	8.5%
T0	Clothe fit	fit, size, clothe, look, weight, big, wear, lbs, small, scale	5.6%

Table 6 - The Result of Topic Modeling (LDA)

1. Weight Management

This category represents the largest theme and discussions in the *loseit* subreddit. Here, we could see that users actively share their personal experiences, challenges, and successes in their weight loss journeys.

T2. Weight change

This is the largest topic within this theme and the entire LDA model. It mainly captures the users' journeys of losing weight. The top keywords, such as "weight" and "lose" could reflect the main focus of the *losit* subreddit on weight loss discussions. Most posts on this topic discuss and encourage others about the challenge of maintaining consistent weight loss over time. As illustrated in the examples below, we can see that in this subreddit, users frequently share their experiences and detail how much weight they have lost over a certain period, whether weekly, monthly, or longer. This is reflected in keywords like "lbs", "pound", "week", "month" and "day". Additionally, some posts involve users seeking advice and new ideas when they hit weight loss plateaus. This is indicated by keywords like "start", "gain" and "loss". Moreover, some other posts share tips and strategies that have helped them overcome these challenges during their weight loss journey.

> *"The first four days I lost 4 pounds, and then so on off lost 8 pounds ever since then. What I'm concerned about is I'm not losing excess weight, for the last 4 days I've been the same weight of 261Ibs, since today checking my weight."*

> *"I did two weeks of Insanity and have seen amazing improvements to my cardio and some muscle tone, but I just weighed in this morning and I weigh THE EXACT SAME as two weeks ago :("*

> *"In the past 6 months, I've lost roughly 25 pounds. That's not nothing. I'll take 25 pounds in 6 months. It's better than the alternative. I've got to stop comparing myself to people who lose 100 pounds in a year or whatever. They aren't on my journey. "*

T11. Weight loss goal

The top keywords such as "goal", "want", and "keep" indicate that this topic focuses on users' personal experience to set and achieve weight loss or fitness goals. Most posts discuss how these goals help them stay on track and overcome challenges on the weight loss journey. For example, some users might set alternative goals like cycling a specific distance to stay motivated. In some other posts, users express the confidence gained from reaching weight goals, such as losing an amount of pounds, as reflected by keywords "weight" and "lbs". Additionally, some posts also celebrate small victories that contribute to long-term success.

> *"I was talking to a friend of mine who is also in the process of losing it and we got to discussing personal goals that weren't weight related… To take some of the pressure and frustration out of weigh-ins, we both gave ourselves goals that we can focus on that will also hopefully have the side benefit of seeing changes on the scale!... On days/weeks where i'm plateauing on the scale I still feel as if i'm being productive because i'm able to track progress towards my cycling goals."*

"Through CICO over the last 3 months, I've lost a total of 32lbs and I'm a little over a third of the way to my goal weight of 135… I wouldn't go the last 2 years because i felt really bad about my body. This year i hope to have the confidence to go!"

"For me, this morning the old man whose wife has the same physical therapy spot as me asked me if I'm a professional athlete. And even though I realise the man has obviously never met a professional athlete, I'm still going to surf that high for as long as I can :D… But also, partly I just want to check in on my old habits, to make sure I'm still on the right track. I don't know if I'll log next week or not. We'll see."

<u>*T9. Medication*</u>

This topic mainly focuses on health-related issues, as discussion within this topic often includes seeking professional medical advice for different health conditions like obesity or other chronic illnesses. As indicated by keywords, "start," "back," and "year", many posts describe the frustration of trying to lose weight but gaining it back. Conversely, some posts share how medical treatment has significantly improved their health, daily lives, and long-term goals. Additionally, keywords like "weight," "gain," "lose," and "lbs" could correspond to the central concern of weight management within this topic.

"… I lose my motivation. I feel like to have such a long journey ahead of me and it's incredibly daunting. I'm not even sure where to start. Every day that I start eating healthy ends up with me eating junk by the end of the day and "starting over again" the following day."

"Fast forward to now and a new specialist I saw gave me medication to help with my overproduction of testosterone (I'm female) and for the first time in forever I've lost 5 kg without gaining it back in a two month period. And I dropped about 1.5 sizes."

"But what I didn't expect was that in a lot of ways my fitness level doesn't seem to change, even when I'm active for prolonged periods and losing weight... I'm currently trying to get an appointment with a dietician, a physical therapist, and, who knows, maybe a personal trainer eventually."

<u>*T1. Daily updates*</u>

This topic discussed the daily routines and activities for weight management. Many posts provide regular updates on a daily, weekly, or monthly basis to record the weight loss process, as reflected in keywords like "day," "week," "today," and "month". As demonstrated in the example post below, we could see that many users share details of a specific day, including what they ate, how they managed their calorie intake, or what exercise they completed.

"Is it Friday yet? Long day today followed by an extremely long day tomorrow. Send me all your good vibes, please? But let's discuss our goals, shall we?!
Diet/Lifestyle Change: GW is 260 by EOM. (Currently sitting at 265 with Shark Week present...let's see how this weekend goes)
Food:Meat is only allowed 4x a month (Still at 02/04)
Food: Snacks! I suck at it, I still love my potatoes (Had blueberries for one snack and tomatoes for the other!)
Draw at least one thing a day (have corporate job work, so probably not tonight)
Dance: Got my soca on yesterday. Today was a rest day and I have another class lined up for

Thursday!
Write: Have a new idea for either my first book or a new tv show. Time to put pen to paper.
Goal: have either a complete outline or rough spec (I'm writing away!!!)
How is your day going?"

2. Diet

This category represents another important aspects of weight loss in the *loseit* subreddit. Key topics within this category include meal planning, calorie tracking, and the challenges of maintaining a nutritious diet.

T10. Emotional eating

This topic mainly addresses the relationship between emotions and eating behaviors. Discussions often include the difficulties of controlling emotional eating and maintaining healthy eating habits. Keywords like "eat," "food," and "binge" could reflect the focus on emotional eating within this topic. Many posts discuss how negative emotions influence their eating behaviors. For example, users share their experiences of turning to food for comfort and overeating during difficult times. Terms such as "calorie," and "hungry" often appear in posts, which indicates that users struggle to balance their emotions with their dietary goals. Additionally, some posts also describe the frustration of maintaining a healthy lifestyle with these emotions, as reflected by the keywords "healthy,""think," and "want".

> *"I ate around 700 calories for lunch, got satisfied and didn't feel hungry anymore. 30 minutes later I kept on eating even though my stomach hurt from being too full. I don't know what happened. 2 hours later now and I've eaten over 3000 calories for no apparent reason. There are times when I'm genuinely hungry and eat 3k kcal, but today I just went haywire just because…"*

> *"Despite being a fairly normal and healthy weight, I 100% have an unhealthy relationship with food. I'll eat absolute garbage, completely let myself go for a month or two, gain 10lbs, then start going super hard, running and lifting every day, cold turkey on junk food, and lose 10lbs… My issue is that I struggle to do anything in moderation… When I do decide to actually be healthy, I'll lift and run 6 days a week and eat SUPER clean. But as soon as I slip up, my body is like ALRIGHT GAME ON, and thus begins another binge period."*

> *"However, when I reached my goal weight I became ravenously hungry and began binging on keto friendly foods. I eventually began binging on non keto foods and gained 75 lbs… I want to lose weight again but I'm afraid that I'll repeat the same pattern. I'm able to lose weight but I have no clue how to maintain it. I don't think I want to be on a ketogenic diet for the rest of my life but it was the only diet that had worked for me."*

T12. Food choice

This topic focuses on the decisions users make about meal planning and nutrition when they are losing weight, as indicated by keywords like "eat," "meal," "food," and "calorie". As illustrated by the example posts below, many discussions within this topic talk about how to manage their dietary needs with different meals such as "breakfast", "lunch" and "dinner". For instance, users often share their strategies for preparing nutritious meals. Terms like "cook" and "chicken" are mentioned frequently in posts, as chicken is a healthy ingredient for a balanced diet. Additionally, some posts often mention " snack " as an unhealthy food when maintaining a healthy diet.

> *"As a contributor I was privy to all the food and encouraged to take as much as I wanted. Breakfast was easy enough. I just stuck to fruit and tea. Lunch was much harder. Thankfully, there was veggies and grilled chicken and I (sadly) walked past the rice and bread. But not dipping my food in the cheese fountain hurt a lot!"*

> *"I decided to go out with some friends last night and have a cheat meal, to get beers and enjoy some of the fried foods I'd been missing. We shared sticky chicken wings with blackened ranch and duck fat frites with garlic aioli. I ordered a soft shell crab po boy and drank a couple ciders… Until I woke up at 6am this morning completely dehydrated and lethargic, and have been sitting on the toilet for 20 minutes now. I forgot how shit all of this food makes me feel."*

> *"I know 12lbs isn't really all that much weight, and I still have a ways to go, but I just feel so powerful right now! For me, I needed to believe in myself first and take care of my mental health… I found low-cal options that I actually liked to eat. Chicken soup from Panera? Yes please I'll have a group serving (410 cals)! Frozen blueberries and fat free cool whip (a boatload for 300 cals)!"*

T3. Calorie tracker

The practice of calorie tracking is the center of this topic, which is crucial to weight loss. Keywords like "calorie," "eat," "deficit," and "burn" emphasize the focus on managing energy intake relative to expenditure. Within this topic, many posts discuss efforts to stay within a specific calorie limit and the impact it has on weight in a certain period, as reflected in terms "day," "week," "weight," and "scale". Some other posts also talk about monitoring their food intake while tracking calories, which makes "eat" a significant keyword. Additionally, Keywords like "fat" and "muscle" appear frequently in these posts when maintaining a consistent calorie deficit over time.

> *"I've seen a lot of people talking about plateauing across the various weight loss/diet related subs I sub to. I've also seen a lot of people saying that weight loss is about pretty simple math, calories in < calories out, and if you're at a deficit, you will lose weight. If the latter is the case, how does the former happen? Is it an issue of not adjusting your calories/deficit to accommodate your TDEE enough?"*

> *"I always count my calories perfectly, I don't add any sauces, oil, nothing. I weigh everything, and I very very rarely cheat with one meal (perhaps once a month)."*

> *"My maintenance level is 2,000 cals, and I have gone as far as eating 1400 calories (which of course is not sustainable but I just did it as an experiment) and not even that works. I am stuck like gum under a table at 176lbs."*

T14. Diet control

This topic captures the challenges of making and sustaining healthier dietary choices. Discussions within this topic mainly talk about how to control sugar intake, beverage choices, and calorie consumption. Keywords like "cut," "sugar," "eat," and "diet" could reflect this topic theme. Notably, "drink" and "water" are the top 2 keywords in this topic, as water is 100% calorie-free and helps reduce sugar and calorie intake. Additionally, terms like "soda" are frequently mentioned in some posts as an unhealthy soft drink to avoid.

> *"Anyway, one of my biggest struggles has been soda. I absolutely love soda, and I could literally drink 1500 calories a day in soda easy. And I've always hated diet soda… But I was so frustrated these past few days that I couldnt fit any soda into my calories, so I broke down and bought some diet (feisty cherry diet coke). I got me a can, and opened it up, and reluctantly tried it- and had to double check that it was actually diet. It was really good! I'm so excited I can finally drink soda again without basically giving up a meal for it!"*

> *"Last June I hit my goal weight of 125 and have been up and down a few pounds since then because life that's why. I always bounce back to the goal weight through cutting just 100 calories under maintenance of I need to… This week I cut down to 1400 just to see if I could short term cut out the trend but it seems it is still climbing… This month has been extra stressful at work so I have been drinking more coffee and sleeping less."*

> *"I'm going to stop drinking I think, more to help with mental health than anything. Or I'll at least limit myself! I usually drink every weekend but I want to stop. I think it could help me with weight as well, I forget how calorie dense it is. It also makes me binge eat! It just drives me crazy that some of my friends all drink a LOT, eat all the McDonald's that night and never seem to gain weight or feel bad over it."*

3. Physical Exercise

This catergory provides an overview of physical activity in the *loseit* subreddit. Within this theme, users share their experiences with staying motivated, exploring different workouts, and using fitness App or technology to tracking and achieve their goals.

<u>*T4. Motivation*</u>

This topic mainly talks about the emotional challenges users face when trying to maintain motivation, as reflected by keywords like "feel," "work," and "hard". Many posts on this topic discussed that it is difficult to stay motivated by regular workouts or healthy eating habits. As the example posts present, users usually share the struggles of maintaining weight and negative emotions throughout the weight loss journey. In some other posts, users find themselves having to restart or regain the cyclical nature of motivation, as suggested by terms like "back," "time," and "start".

> *"My weight has been climbing, and I'm trying to do what I did before hitting my goal weight...but I seem to be screwing up every other day. It's so odd that I know what to do to lose weight but can't seem to maintain at all...I think about food in all my hours awake, and it's honestly very draining. How do I stop self-sabotaging? "*

> *"I'm struggling lately to not give into my little cravings and stay motivated to work out even though I'm so happy thus far with my results. I feel like that should be enough reason for me to keep working hard and eating right, but...I feel a little bit of bitterness and resentment towards my situation because now I feel like I can't really "enjoy" myself anymore."*

"I've wanted to lose weight for many years but as someone who's battled anxiety and depression, even thinking about it felt stressful. Especially when depression had depleted my motivation...In order to stay motivated, I have to clean up other aspects of my life...This was the motivation I needed."

T15. Workout plan

This topic focuses on how to establish and maintain effective workout routines. Discussions often include starting new exercise regimens or exploring more workout types, as reflected by keywords like "gym," "workout," "exercise," and "routine". Within this topic, many posts share workout plans, the challenges, or the successes they face and achieve in weight loss. As the example post presents, some users start to take fitness or advanced classes such as body pump and HIIT, as indicated by the terms "cardio" and "work". Additionally, other users ask for more enjoyable exercise to avoid boldness when maintaining a regular exercise activity.

"I feel like I'm ready for the next step which would be the gym for me. I've already gone to (and struggled at) a Yoga class. I had an assessment at the gym where they made a report of how fit I am (testing cardio, physical, measurements, etc) and I was given a plan to start getting back into fitness which is a combination of Spinning, BodyPump, CxWorx, and Yoga. I have my first BodyPump class tomorrow morning and am so unbelievably nervous."

"Started doing the general 10k a day and have since worked up to 20-45k a day (minus Sunday, cheat day)… The beginning of this year I started kickboxing & HIIT classes & try to get in some Fitness Boxing on the Switch when I don't have time for class… I feel amazing and so proud of myself today. I cannot wait to see what the next year looks like as I build more muscle and continue on this journey."

"When I go to the gym, I only run on the treadmill. It's starting to get boring and repetitive. I want to find more fun ways to burn calories and exercise without having to go to the gym all the time or using the treadmill. I don't know how to use the other machines at the gym either."

T5. Fitness App

This topic focuses on using technology to monitor and manage fitness metrics. The fitness technology mainly points to fitness apps or tracking tools, as indicated by keywords like "calorie," "us," "app," and "track." Many posts discuss using apps to track calorie intake, daily activities, and other fitness metrics (e.g., heart rate) to achieve their body and weight goals. Terms such as "daily," "body," and "fat" suggest that these apps or tools are used consistently to manage weight loss outcomes. In some posts, users also discuss how to choose the suitable fitness device, while others share their experience with a premium app membership to fine-tune calorie tracking.

"I used my 'fun' money to buy myself a food scale and a LoseIt premium membership to fine-tune my tracking...The scale has been moving, but I'm glad I now have one more tool to keep track of the calories I'm taking in. I have a feeling that would've gotten extremely problematic later on."

"I am looking to upgrade to either a new watch or a Fitbit...I guess I am curious about anyone's experience with the Versa, or Fitbit in general. Only thing I'm truly concerned about is data tracking and how much is stored on their end about my biometrics and how it might potentially hurt me if things got hacked."

"Her heart rate is exceptionally low (and it's not a disease) as a result of her workouts. At resting she is at about 40 bpm. When she takes spinning classes her heart rate is always the lowest of the class (it's publicly displayed). When we go for a run together, right after the run she is immediately back at like 60 bpm when I'm still at 140 bpm."

T6. Exercise routine

This topic captures both the challenges and rewards of maintaining an exercise regimen. Many posts showcase personal achievements such as completing a long run and sticking to a workout schedule, as suggested by keywords "run" and "walk." Some posts also discuss running distances, walking routines, and the time dedicated to these activities, as reflected by keywords such as "minute," "mile," "week," "day," and "time".

"Yesterday night, I ran 3.5 miles outside without stopping, the first time I've run that long of a distance since 2017... It didn't even feel hard, it just felt like being on autopilot and like I could go on for a long time, which wasn't the case in the past."

"I made it! Went to the gym, did 30 minutes on a stationary bike and the ab circuit... It wasn't much, but it got my blood pumping and confirmed exactly what everyone has said: getting there was the hardest part."

"I lost 10 pounds and it's thanks to calorie counting, 30 minutes of cardio and as well as doing good meal preps for 2-3 days... To keep myself entertained while doing cardio, I usually watch 1-2 chapters of an anime."

4. Emotions and Support

This category focuses on mental and emotion problems and support related to weight loss progress in the *loseit* subreddit. Within this theme, users often seek advice and support for their struggles in relationships and weight loss efforts throughout their journey.

T13. Negavtive feelings

This topic discusses the emotional and mental challenges users face in their personal relationships when keeping a healthy lifestyle, as indicated by keywords like "life," "people," "health," "family," and "friend". Discussion within this topic often involves internal struggles and feelings about their relationships. For example, one post shares discomfort and guilt feelings when refusing a friend's invitation. Another post expresses concern about his wife's dissatisfaction with her body and the strain on their relationship. Additionally, the third post share the exerperience and inspiration from a doctor that addressing the underlying emotional reasons for overeating was key to achieving sustainable weight loss.

"I've been doing CICO for a few weeks now and have been feeling great! However I am very careful with what I eat which doesn't lend itself well to social situations, no entrees, no alcohol, no to most food… Due to this I keep declining offers, which makes me a horrible friend. I would have no qualms telling this friend about my diet and reasons why I can't hang out, except for this friends background. She's very opinionated and was in the hospital for several months as a youth for anorexia?"

> *"Every time I look at my wife, I fall in love with her all over again, from her point of view though, she really, really hates her body... The problem is that what motivates me demotivates her... I don't care how much she weighs, I don't care what she looks like, I just want her to be happy with herself. I'm really in a tight spot because if I push her too much, she'll think 'my husband thinks I'm fat.'"*

> *"I had decided 2019 would be the year I'd lose weight "for good" but as it turns out, I was randomly diagnosed with chronic kidney disease in January which turned everything around… He was interested in WHY I over-ate rather than WHAT I should eat. All in all, he explained that if I did not develop healthier strategies to COPE with life (stress, whether big or small), all weight loss efforts would ultimately end up with me re-gaining the weight."*

T8. Seeking advice

This topic captures discussion on seeking and sharing advice related to weight loss, as indicated by keywords like "help," "advice," and "need". Terms such as "diet," "weight," and "lose" reflect that the users mainly seek advice about managing weight and optimizing dietary habits in this subreddit. However, this advice varies widely due to the large amount of users and their diverse experiences. For example, one post shares the experience of seeking advice from a Facebook group while trying to find more reliable and accurate information on weight loss. Another post asks for more details about the safety of extreme calorie restriction over the short term. Notably, "thank" is also an important term in this topic. As presented in the third example, many posts express gratitude for the advice and support received from this subreddit.

> *"Hi, I belong to a Facebook group for losing weight with saxenda. I really see how crazy it can be taking advice in groups like this… I've been looking for good articles on the mechanics and truth of starvation mode, googled a lot, does anyone have and good solid recs for information sources on this?"*

> *"I'm asking for help Here's my first question. Would cutting down to 600-800 calories for the last 9 days (final weigh in is next Thursday) cause any long-term issues? From what I've seen, crash dieting like this is only really bad if you do it long-term. How low can you go with that for a short-term cut like this?"*

> *"Thank you! For some reason Im banned from commenting right now but thank you everyone for the amazing feedback, and to those similar to me who found some inspiration in this post, Im so glad you could, and I wish you all the same experience I've had! Do not give up if this is your goal, it CAN be done."*

5. Appearance

This category mainly focuses on body image and clothing fit during weight loss. It is also a very important theme in the *loseit* subreddit as weight loss outcome is often directly associated with changes in body appearance at first glance.

T7. Body image

This topic explores the struggles users face with body image, as suggested by keywords like "look," "body," "weight," and "fat". Discussions often involve how users perceive their bodies and the emotional impact of these perceptions, as reflected by terms like "feel," "want," "know," and "think". Many posts on this topic expressed dissatisfaction with not meeting personal or societal expectations of body image. As presented in the example posts, some users even share horrible experiences of getting bullied or ridiculed in their daily lives. Others share their experience about focusing on their health and self-acceptance over conforming to societal standards such as being skinny.

> *"I finally worked up the courage to go to the gym and on my second day I was doing the chest press machine. The machine faces a mirror with other machines on each side and I look over and some kid about my age is recording me. I tried to ignore him but he literally sat there for 10 whole minutes just staring and laughing at me. Since then I haven't gone back."*

> *"I personally feel no desire to be "skinny", I just want to be healthy and feel good in my body. I've seen people aim for the top of their "ideal" range, the middle, the bottom, or even outside of it, and I'm curious about how/why you decided on the goal weight/window you did or, if you haven't yet, what method of choosing you plan to use."*

> *"On March 1st, I decided that I didn't want to live my entire life hating how I looked, so I seriously began my weight loss journey... I want to go to college with a true sense of confidence, and be able to explore all of the opportunities I denied myself because of my insecurities from being overweight."*

<u>T0. Clothe fit</u>

This topic discuss the significant role that clothing plays in weight loss, as indicated by the keywords like "fit," "size," "clothe," and "look". For many people, the way their clothes fit serves as an indicator of their progress in weight loss. This is reflected to terms "big" and "small". As the example posts show, some posts express positive feelings like pride and joy when they notice their clothes fitting better or moving down to smaller sizes.

> *"So finally I get a pair of 36s on without struggling and for the first time in my life I knew what it felt like for clothes to fit right. I genuinely got emotional seeing the fruits of my labor coming together."*

> *"I went to a few stores and after trying on a couple sizes I realized I was able to fit into the REGULAR sized section. I'm not an XXL or XL anymore I'm a L ! I even bought a jacket that was a medium!"*

> *"But today, almost two weeks into my journey, it closed!! I was able to zip it up! Not only did it close with no effort but I'm only a pound or two away from closing and not looking like maybe it's a tiny bit too tight. It's a good feeling."*

4.1.2 Topic trend analysis

After interpreting each topic's keywords and contents generating from LDA moel, I would also like to see how these discussions have changed over time. In this section, I visualized the topic trends in five catergories which were grouped in the previous section 4.1.1. Figure 11 to 15 below illustrates the changing proportions of each topic within these themes in the *loseit* subreddit from March 2019 to March 2021. Moreover, dots on the charts represent the monthly proportion of each topic and different colors correspond to each specific topic.

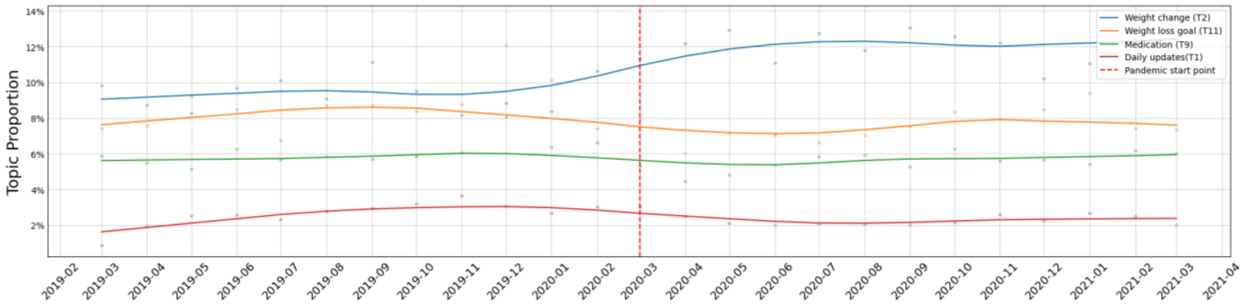

Figure 11 - Monthly posts proportion of topics in the Weight Management Theme (T2, T11, T9, T1)

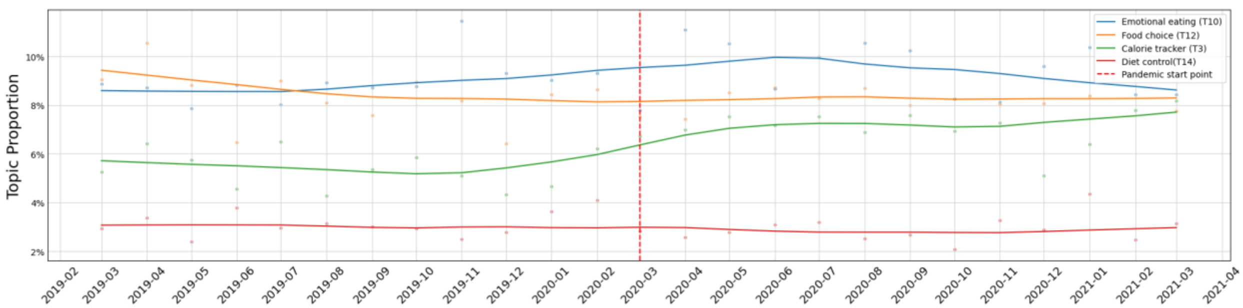

Figure 12 - Monthly posts proportion of topics in the Diet Theme (T7, T8, T13, T14)

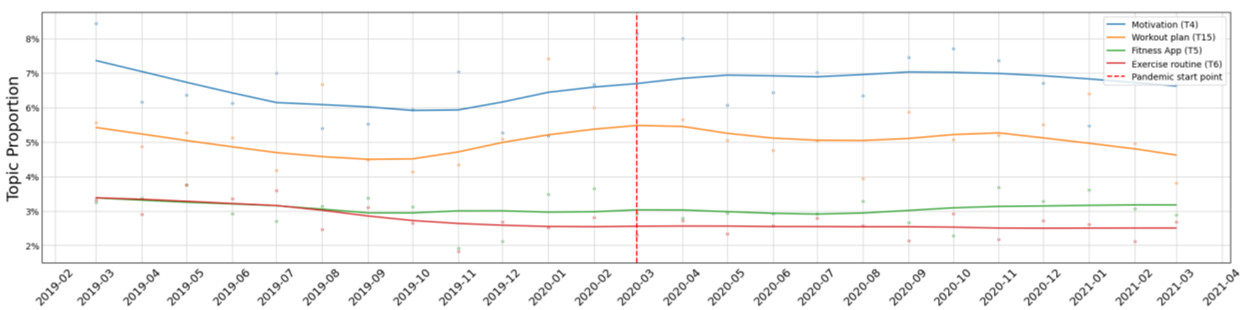

Figure 13 - Monthly posts proportion of topics in the Physical Exercise Theme (T4, T15, T5, T6)

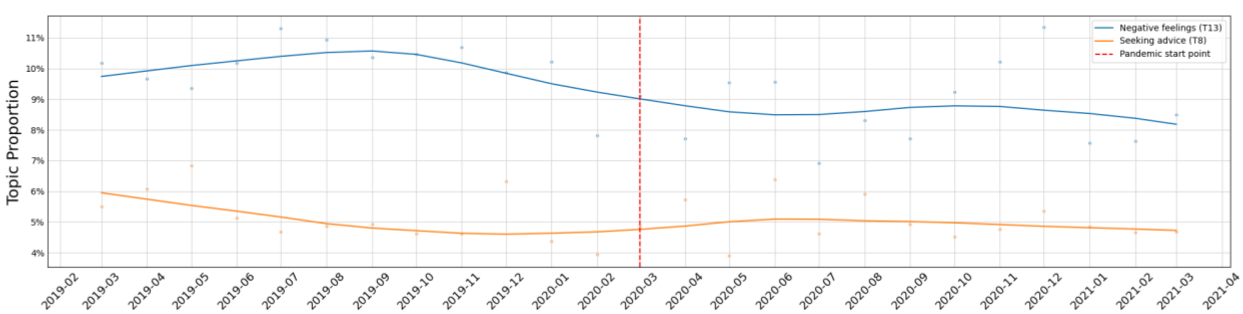

Figure 14 - Monthly posts proportion of topics in the Emotions and Support Theme (T13, T8)

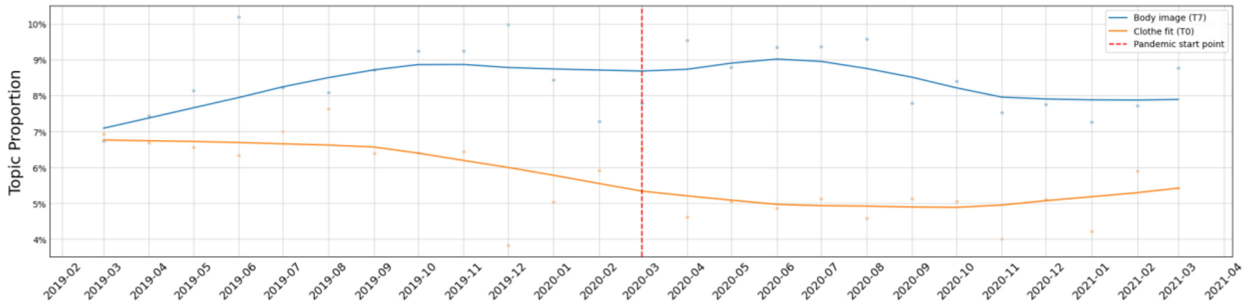

Figure 15 - Monthly posts proportion of topics in the Appearance Theme (T7, T0)

As Figures 11 to 15 show, we found several topics change significantly throughout the study period in the *loseit* subreddit. In the Weight Management theme (Figure 11), there was an increased focus on *Weight change* (T2) starting at the end of 2019 and continuing untill the pandemic outbreak. In the Diet theme (Figure 12), *Calorie tracker* (T3) showed a significant increase over time, especially in the months following the outbreak. *Emotional eating* (T10) fluctuated slightly, with a small increase during the early month of the outbreak. In the Physical Exercise theme (Figure 13), both *Motivation* (T4) and *Workout Plan* (T15) showed some fluctuations with increasing trends during the outbreak and then stayed stable. In the Emotions and Support theme (Figure 14), *Negative feelings* (T13) showed a clear fluctuation and declining trend over time. In the Appearance theme (Figure 15), *Body image* (T7) increased while *Clothe fit* (T0) declined before the outbreak, and then both topics maintained relatively stable trends during the pandemic. Other topics, such as *Weight loss goal* (T11), *Medication* (T9), *Daily updates* (T1), *Food choice* (T12) and *Diet control* (T14), *Fitness App* (T5), *Exercise routine* (T6) and *Seeking acvice* (T8) remained relatively stable and slight fluctuations over time.

4.2 Results of Sentiment analysis

4.2.1 Sentiment polarity

Following the method described in Section 3.4, I present the results of the sentiment polarity analysis by VADER lexicon tools in this section. The results includes the overall sentiment trends and sentiment polarity within different topics in the *loseit* subreddit.

	Pre-pandemic		**Dur-pandemic**	
Sentiment	***Num. of Posts***	***Percentage***	***Num. of Posts***	***Percentage***
Positive	11,455	70.53%	13,294	69.46%
Negative	4,572	28.15%	5,619	29.36%
Neutral	215	1.32%	226	1.18%

Table 7 - Number of posts and proportions of sentiments (positive, negative, netural) across two time periods

Table 7 is a descriptive table for the number of posts, and proportions of sentiments in posts for each period. As indicated by Table 7, positive posts outweighed significantly than negative posts during both periods.

However, during the pandemic, positive posts decreased slightly by 1.07%, while negative posts increased slightly by 1.21%. Additionally, the neutral posts changed little (+0.26%).

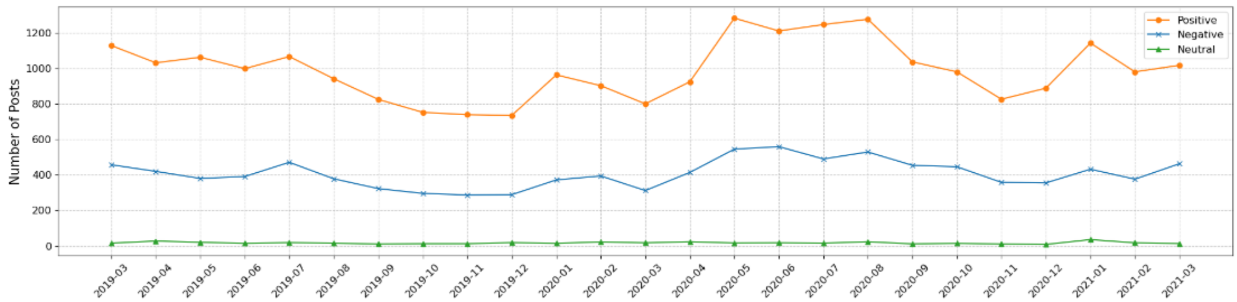

Figure 16 - Number of posts of sentiments (positive, negative, netural) trends over time

To explore the sentiment trends, I visualized how sentiment evolved from March 2019 to March 2021, As illustrated in Figure 16. The sentiments corresponded to the result in Table 7, showing that there are more positive posts than negative ones over time in this weight loss subreddit. However, from March 2020 to May 2020, during the early time of the pandemic outbreak, we saw increasing trends in both positive and negative posts about weight loss. Moreover, May 2020 was marked as a distinct peak during the pandemic. It was sustained through August before declining again. This suggests that the users discussed more weight loss-related topics during this period. Notably, two notable peaks were also observed in January 2020 and January 2021, which reflects a seasonal pattern of increased activity. This could potentially relate to the New Year resolutions, that people might be more motivated to set weight loss goals at the beginning of the year. Additionally, in this subreddit, neutral posts remained stable, with only slight changes over time.

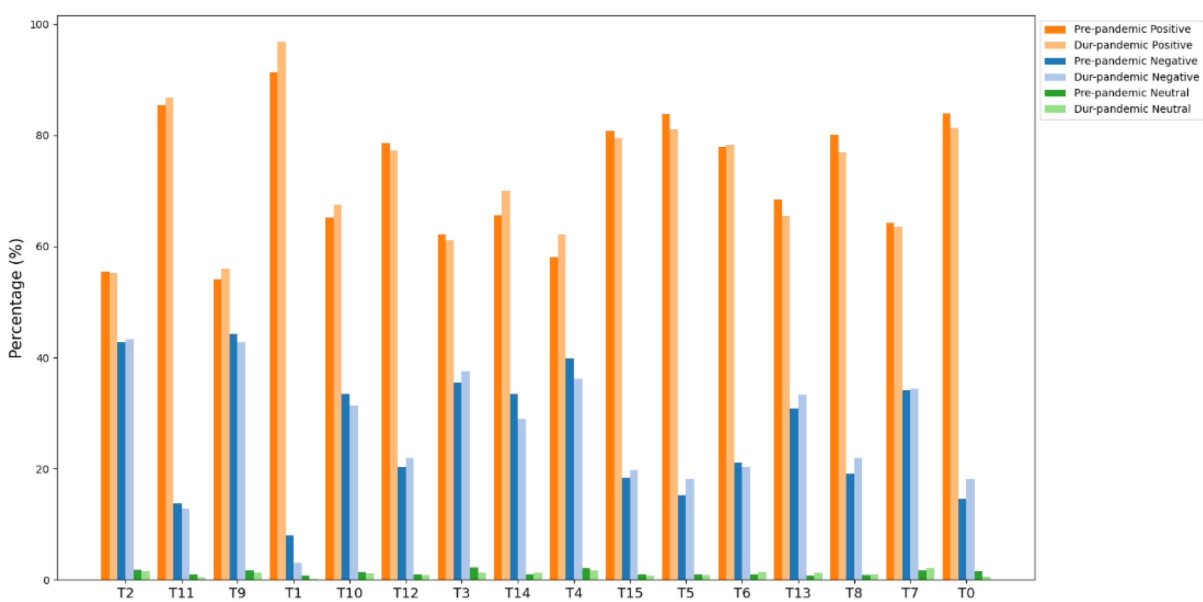

Figure 17 - Proportions of sentiments polarity (positive, negative, netual) by topics across two time periods

After analyzing the overall sentiment polarity trends by VADER, I also examined the sentiment proportions at the topic level to explore how positive, negative, and neutral sentiments shifted in specific topics over time. Figure 17 below presents the proportions of sentiment polarity (positive, negative, neutral) for each topic before and during the pandemic. The detailed table supporting Figure 17 can be found in the appendix tables.

From Figure 17, we could observe several significant sentiment changes in topics during the pandemic. In the Weight Management theme (T2, T11, T9, T1), positive posts in *Daily updates* (T1) increased (+5.42%), while negative posts decreased during the pandemic (-4.93%). In the Diet theme (T10, T12, T3, T14), posts about *Emotional eating (T10)* saw an increase in positive sentiment (+2.28%) with a decrease in negative sentiment during the pandemic (-2.04%). Furthermore, positive posts in *Diet control* (T14) also saw an increase (+4.30%), while negative posts decreased (-4.55%). In the Physical Exercise theme (T4, T15, T5, T6), positive posts (-2.70%) in *Fitness App* (T4) also decreased while negative posts increased (-3.99%). Conversely, positive posts (+3.99%) in *Motivation* (T4) also rose while negative posts decreased (-3.63%). In the Emotions and Support theme (T13, T8), positive posts in *Negative feelings* (T13) decreased slightly (-3.00%) while negative sentiments increased (+2.47%). Moreover, positive posts in *Seeking advice* (T8) decreased (-3.11%) during the pandemic, with an increase in negative posts (+2.90%). Additionally, in the Appearance theme (T7, T0), positive posts in *Clothe fit* (T0) declined slightly (-2.65%), and negative posts saw a corresponding increase (+3.56%) during the pandemic.

4.2.2 Emotion distribution

This section presents a deeper emotion analysis using NRC-EIL lexicon tools. This includes analysis of the overall emotion trends and the distribution of emotions within specific topics in the *loseit* subreddit. First, we look at the descriptive statistics of the emotions in this weight loss subreddit. Table 8 below presents the frequency and proportion of words related to eight emotions in posts before and during the pandemic.

	Pre-pandemic		**Dur-pandemic**	
Emotion	***Words***	***Percentage***	***Words***	***Percentage***
Joy	140,532	12.96%	163,295	12.85%
Trust	165,270	15.25%	191,707	15.08%
Anticipation	191,040	17.62%	223,038	17.55%
Surprise	106,091	9.79%	124,785	9.82%
Fear	141,028	13.01%	169,208	13.31%
Sadness	138,341	12.76%	163,727	12.88%
Disgust	110,334	10.18%	130,163	10.24%
Anger	91,312	8.43%	105,152	8.27%
Total	1,083,948	100%	1,271,075	100%

Table 8 - Frequency and proportion of words associated with eight emotions across two time periods

As shown by Table 8, a total NRC-EIL word count of 2,355,023 with 1,083,948 words before and 1,271,075 words during the pandemic was exhibited in this weight loss subreddit. Specifically, positive emotion words such as *joy* (-0.11%), *trust* (-0.17%), and *anticipation* (-0.07%) decreased slightly during the pandemic. This corresponded to the overall decrease amount in positive posts in Table 5. On the other hand, positive emotion words such as *fear* (+0.30%), *sadness* (+0.12%), and *disgust* (+0.06%) showed a slight increase trend during the pandemic. This correlated to the rise in negative posts observed from the same period in

Table 5. Additionally, *surprise* (+0.03%) and *anger* (+0.04%) increased little during the pandemic. Although these changes are small, given the total word count, this could reflect the significant differences in the emotions about weight loss discussions during the pandemic.

To further explore the emotion trends in the *loseit* subreddit, I visualized the proportion of emotion words over time, as illustrated in Figure 18. Although the proportion of each emotion word did not change significantly and remained within 1%, we could still observe trends of each paired emotion.

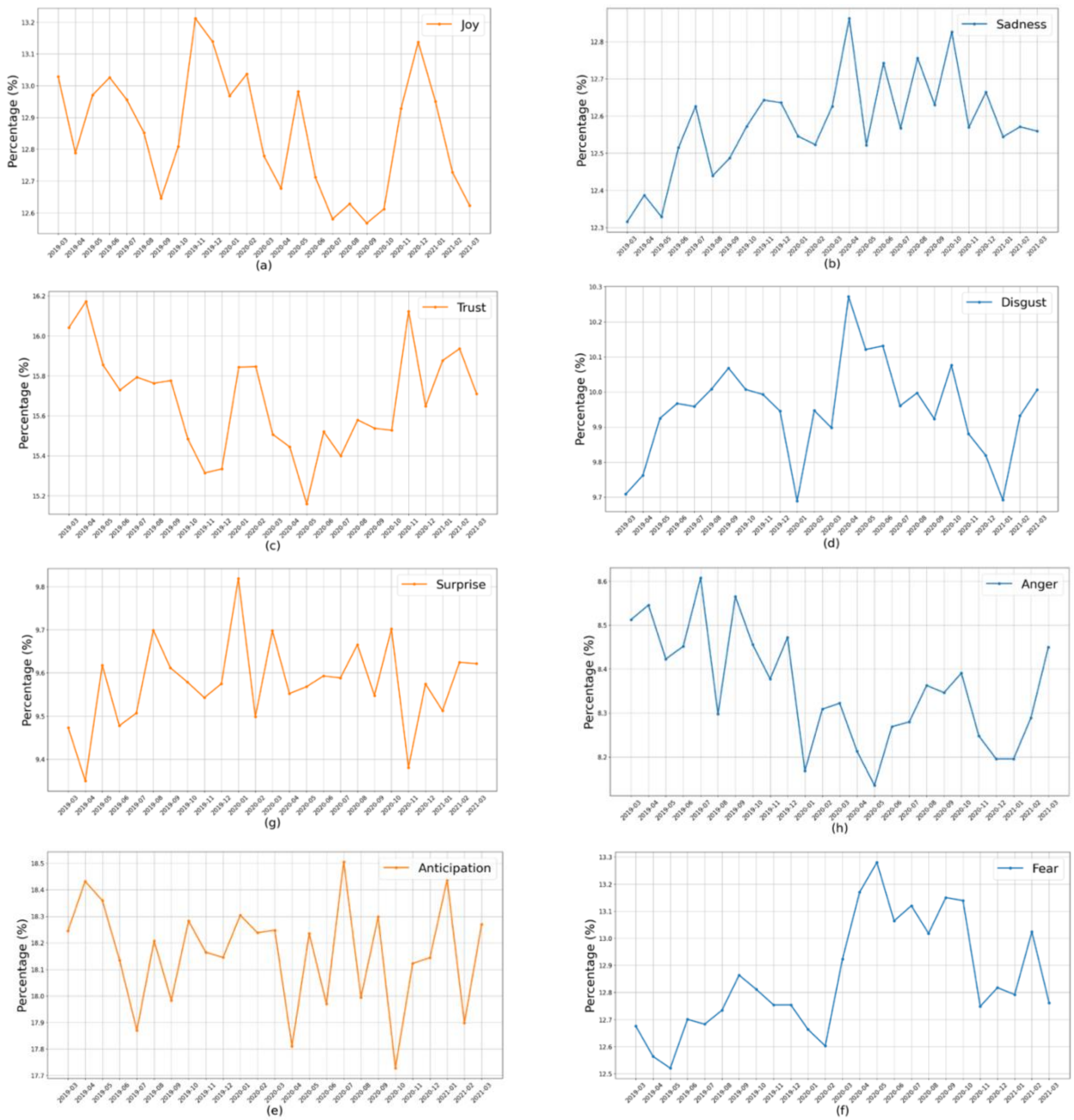

Figure 18 - Monthly Proportion of emotion words counts over time

Joy (18a) and Sadness (18b): Emotion words related to *joy* fluctuated between 12.6% and 13.2% from March 2019 to March 2021. Although *joy* slightly declined at the start of the pandemic, it slowly recovered from May 2020 onward. Conversely, *sadness* fluctuated slightly, ranging from 12.3% to 13.2%. Notable increases occurred in April and October 2020, possibly due to the pandemic's impact and seasonal factors. Although *sadness* declined after November 2020, it remained relatively high.

Trust (18c) and Disgust (18d): *trust* experienced fluctuations between 15.1% and 16.1% from March 2019 to March 2021. At the pandemic's onset, there was a slight decline in *trust*. And then, it increased after June 2020, reaching a peak in November 2020. Conversely, *disgust* rose sharply at the beginning of the pandemic, with a notable peak at 10.27% in April 2020. This was the highest point during the period before it decreased afterward.

Surprise (18e) and Anger (18f): Throughout the pandemic, *surprise* fluctuated between 9.3% and 9.8%, with no significant trends. However, we observed two low points in April 2019 (9.35%) and November 2020 (9.38%). Conversely, *anger* fluctuated more than *surprise*. There was a noticeable decrease in *anger* at the beginning of the pandemic compared to the pre-pandemic period. However, *anger* began to rise sharply by the end of 2020 and into January 2021.

Anticipation (18g) and Fear (18h): *anticipation* exhibited significant fluctuations between 17.7% and 18.5%, with a peak reaching in July 2020 (18.50%). Notably, in the middle to late 2020, *anticipation* reached several highs and lows. In contrast, *fear* increased significantly at the start of the pandemic in March 2020, peaking at around 13.2%. Although it declined slightly afterward, *fear* remained high throughout the outbreak.

After analyzing the emotion distribution trends by NRC-EIL, to further gain insight into which discussion got significant change with emotions, I also examined the emotion distribution at the topic level to conduct a detailed analysis of the eight emotions and their proportion in each topic before and during the pandemic. Figures 19 and 20 show the percentage of words associated with eight emotions by topic for each period. Details about these figures are in the appendix. The detailed data supporting Figures 19 and 20 can be found in the appendix tables.

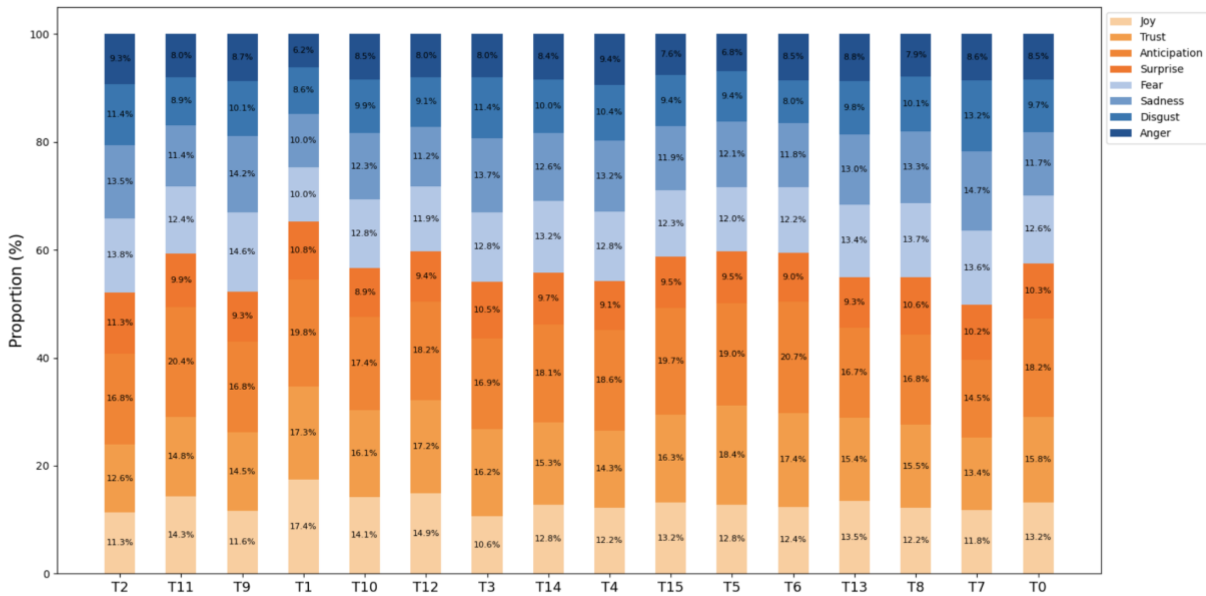

Figure 19 - Proportion of words associated with eight emotions by topics before the pandemic

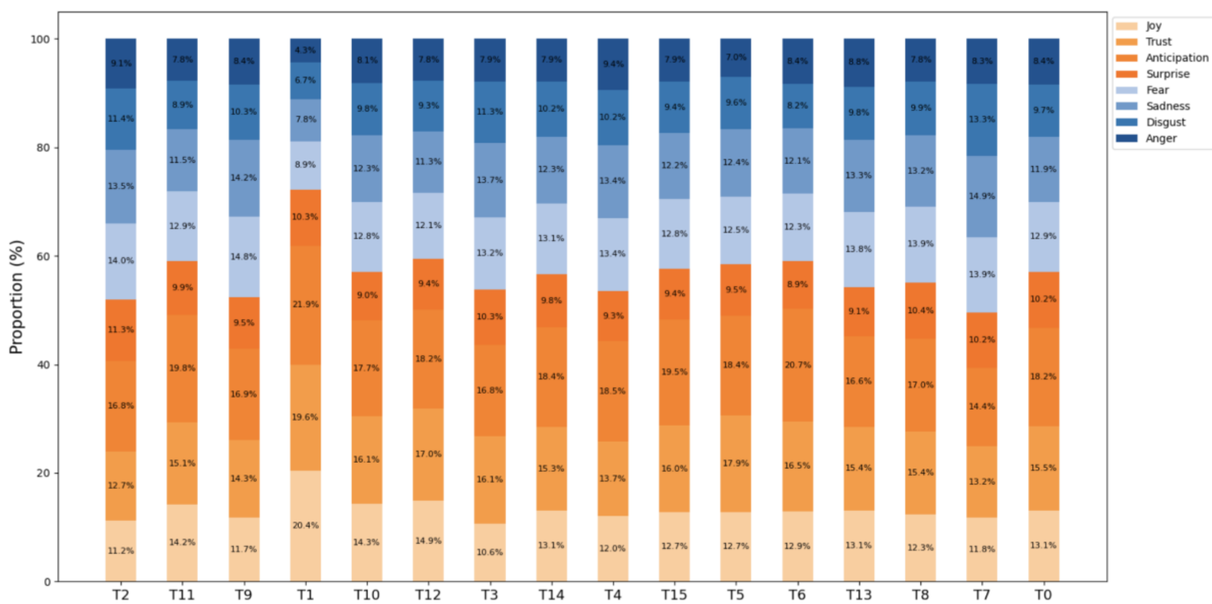

Figure 20 - Proportion of words associated with eight emotions by topics during the pandemic

In the Weight management theme (T2, T11, T9, T1), the emotional changes in *Daily updates* (T1) were the most significant. Among them, positive words such as *joy* (+2.96%) and *trust* (+2.33%) increased significantly in the post, while negative words such as *fear* (-1.09%), *sadness* (-2.16%) and *anger* (-1.82%) decreased significantly in the post. Positive and negative words in *Weight change* (T2), *Weight loss goal* (T11), and *Medication* (T9) changed slightly, and the overall change was not noticeable.

In the Diet theme (T10, T12, T3, T14), positive and negative words in the four topics, *Emotional eating* (T10), *Food choice* (T12), *Calorie tracker* (T3), and *Diet control* (T14) changed slightly, and the overall emotion was relatively stable.

In the Physical Exercise theme (T4, T15, T5, T6), positive words in *Motivation* (T4) decreased, such as *joy* (-1.14%) and *trust* (-0.61%), decreased in the posts. Positive words in *Workout plan* (T15) decreased with *joy* words decreased significantly. Negative words changed slightly in the above topics. In addition, positive and negative words in *Fitness App* (T5) and *Exercise routine* (T6) were relatively balanced in the post, and the change was not significant.

In the Emotions and Support theme (T13, T8), the emotional expression of *Negative feelings* (T13) decreased slightly, with positive words decreasing slightly. Negative emotional words in this topic also changed slightly, *fear* (+0.41%) and *sadness* (+0.29%) increased slightly. The emotions in *Seeking advice* (T8) changed slightly, and the overall emotional expression in this topic was relatively stable.

In the Appearance theme (T7, T0), positive words in both topics decreased slightly, while negative words increased slightly. Positive words such as *trust* (-0.20%) in *Body image* (T7) decreased slightly. Negative words remained stable, with a slight increase, but the change was little. Positive words in *Clothe fit* (T0) also decreased with the *joy* (-0.18%) and *trust* (-0.29%) words decreased. Negative words increased slightly but there was little change.

4.2.3 Analytical results

In order to compare the difference of positive and negative posts in each topic before and during the pandemic, I used the two-proportion z-test for comparison. Tables 9 to 13 show the z-test results based on five categories.

	Positive		**Negative**	
Topic	***z-statistic***	***p-value***	***z-statistic***	***p-value***
Weight change (T2)	0.513	0.608	-1.149	0.251
Weight loss goal (T11)	-3.711	<0.001	2.362	0.018
Medication (T9)	-3.567	<0.001	2.825	0.005
Daily updates (T1)	-21.868	<0.001	20.683	<0.001

Table 9 - Two proportion z-test of postive and negative sentiments by topics in the Weight Management Theme across two time periods

In the Weight Management theme (Table 9), we can see that the sentiment changes in *Daily updates* (T1) were the most significant, with positive posts decreasing significantly ($z = -21.868$, $p < 0.001$) and negative posts increasing significantly ($z = 20.683$, $p < 0.001$). In *Medication* (T9), positive posts also decreased significantly ($z = -3.567$, $p < 0.001$) and negative posts showed a significant rise ($z = 2.825$, $p = 0.005$). Similarly, positive posts declined significantly ($z = -3.711$, $p < 0.001$) and negative posts increased ($z = 2.362$, $p = 0.018$) in *Weight loss goal* (T11). However, *Weight change* (T2) did not change.

	Positive		Negative	
Topic	***z-statistic***	***p-value***	***z-statistic***	***p-value***
Emotional eating (T10)	-4.531	<0.001	4.089	<0.001
Food choice (T12)	3.155	0.002	-3.668	<0.001
Calorie tracker (T3)	2.125	0.034	-4.008	<0.001
Diet control (T14)	-8.646	<0.001	9.227	<0.001

Table 10 - Two proportion z-test of postive and negative sentiments by topics in the Diet Theme across two time periods

In the Diet theme (Table 10), we could see all four topics changed significantly during the pandemic. For *Emotional eating* (T10) and *Diet control* (T14), positive posts decreased significantly ($z = -4.531$, $p < 0.001$; $z = -8.646$, $p < 0.001$), while negative posts increased significantly ($z = 4.089$, $p < 0.001$; $z = 9.227$, $p < 0.001$). In contrast, positive posts in *Food choice* (T12) and *Calorie tracker* (T14) experienced significant increases ($z = 3.155$, $p = 0.002$; $z = 2.125$, $p = 0.034$) with significant decreases in negative posts ($z = -3.668$, $p < 0.001$; $z = -4.008$, $p < 0.001$).

	Positive		Negative	
Topic	***z-statistic***	***p-value***	***z-statistic***	***p-value***
Motivation (T4)	-7.638	<0.001	7.008	<0.001
Workout plan (T15)	2.901	0.004	-3.471	<0.001
Fitness App (T5)	6.633	<0.001	-7.130	<0.001
Exercise routine (T6)	-0.884	0.377	1.842	0.066

Table 11 - Two proportion z-test of postive and negative sentiments by topics in the Physical Exercise Theme across two time periods

In the Physical Exercise theme (Table 11), these three topics *Motivation* (T4), *Workout plan* (T15), and *Fitness App* (T5) showed opposite trends in sentiment change. T4 showed significant sentiment changes, with positive posts significantly reduced ($z = -7.638$, $p < 0.001$) and negative posts significantly increased ($z = 7.008$, $p < 0.001$). In contrast, the other two topics T5 and T15 showed an significant increase trend of positive posts ($z = 2.901$, $p = 0.004$; $z = 6.633$, $p < 0.001$) and a decreased of negative posts ($z = -3.471$, $p < 0.001$; $z = -7.130$, $p < 0.001$). Additionally, the changes in *Exercise routine* (T6) was not significant.

	Positive		Negative	
Topic	***z-statistic***	***p-value***	***z-statistic***	***p-value***
Negative Feelings (T13)	5.985	<0.001	-4.959	<0.001
Seeking advice (T8)	7.096	<0.001	-6.721	<0.001

Table 12 - Two proportion z-test of postive and negative sentiments by topics in the Emotions and Support Theme across two time periods

In the Emotions and Support theme (Table 12), the two topics, *Negative feelings* (T13) and *Seeking advice* (T8) exhibited similar trends during the pandemic. Both of their positive posts saw a significant increase (z

= 5.985, $p < 0.001$; $z = 7.096$, $p < 0.001$) with a significant decrease in negative posts ($z = -4.959$, $p < 0.001$; $z = -6.721$, $p < 0.001$).

	Positive		**Negative**	
Topic	***z-statistic***	***p-value***	***z-statistic***	***p-value***
Body image (T7)	1.281	0.200	-0.386	0.699
Clothe fit (T0)	6.538	<0.001	-9.009	<0.001

Table 13 - Two proportion z-test of postive and negative sentiments by topics in the Appearance Theme across two time periods

In the Appearance theme (Table 13), *Clothe fit* (T0) showed significant sentiment changes, with positive posts increasing ($z = 6.538$, $p < 0.001$) and negative posts decreasing significantly ($z = -9.009$, $p < 0.001$). Conversely, in *Body image* (T7), changes in both positive and negative posts were not significant ($z = 1.281$, $p = 0.200$; z = -, $p < 0.001$).

5.Discussion

Findings of topic modeling

In this study, I analyzed the discussions and the sentiments within the weight loss subreddit – *loseit* to explore how weight loss discussions and sentiments changed on Reddit during the COVID-19 pandemic. I used the LDA topic modeling and trend analysis to address the first research question: *What discussions about weight loss are mainly talked about on Reddit? How did the thematic content of weight loss discussions change from the pre-pandemic period to the during-pandemic era on Reddit?*

To answer this question, I first identified the topics discussed about weight loss on Reddit between Mar 2019 and Mar 2021. As the result of the LDA model shown in Section 4.1.1, we find that there are five themes as the main area focus of weight loss discussions, which are Weight Management, Diet, Physical Exercise, Emotions and support, and Appearance. From these themes, we could see that in this subreddit, users are not only interested in the weight loss discussion (e.g. weight change, calorie tracker) but also discussions related to upstream topics (e.g. weight loss goal, motivation) and downstream topics (e.g. body image, clothe fit). This finding is similar to the previous study in the same subreddit (Yang et al., 2023). However, among these topics, some topics were discussed frequently during the study period, such as *Weight change* (T2), *Emotional eating* (T10), *Negative feelings* (T13), *Body image* (T7) and *Weight loss goal* (T11). In contrast, some topics such as *Daily updates* (T1), *Exercise routine* (T6), *Diet control* (T14) and *Fitness App* (T7) were discussed less often during the study period.

Within these themes, the discussion volumes of specific topics varied over time, particularly as the pandemic progressed. Notably, serval topics such as *Weight change* (T2), *Calorie tracker* (T3), *Motivation* (T4) and *Workout plan* (T15), all showed an increase in discussion amounts. This indicates that users became more concerned with weight loss behaviour, including tracking their weight and caloric intake, maintaining motivation, and planning workouts during the pandemic. However, both *Negative feelings* (T13) and *Clothe fit* (T0) experienced declining trends over time. Although negative feelings were widely discussed in previous research during the pandemic (Low et al., 2020), it is surprising that users discussed less about negative feelings in this weight loss subreddit. This may be because subreddits on Reddit play important roles in providing social support for netizens. As an online virtual community, the *loseit* subreddit could contribute to reduce the discussions about negative feelings. Similar findings were also found in the studies in the same subreddit (Pappa et al., 2017; Cunha et al., 2017). Notably, although the discussions about negative feelings were reduced, we observed a slight increase trend in *Emotional eating* (T10). This means that users discussed emotional eating behaviours more than before. This topic was also widely explored in previous research during the pandemic (Bumatowska et al., 2022; Cecchetto et al., 2021). Furthermore, discussions about the *Clothe fit* (T0) became less. This may be because people had fewer opportunities to go out and try on clothes in the shops during the pandemic. Furthermore, users discussed about *Body image* (T7) more, which suggests that users cared more about their body image during the pandemic. Previous studies have also evidenced that body satisfcation is one of the important indicators of weight loss outcomes (Palmeira, 2010).

Findings of sentiment analysis

To answer the second research question: *How did sentiment and emotional expressions of weight loss discussions shift across these two periods?* I applied both the VADER and NRC-EIL lexicon-based approach for analysis.

First, I analyzed the sentiment polarity and emotion distribution in this weight loss subreddit. VADER and NRC-EIL analysis had similar results in the overall sentiment polarity and emotion distribution. Specifically, the VADER result shows that positive posts consistently outweighed negative posts on weight loss discussions throughout the study period. This indicates that users generally maintain a positive attitude towards weight loss and healthy lifestyles, even during the challenges posed by the pandemic. However, there was a slight but noticeable decline in the positive posts over time, with a slight increase in negative posts. Furthermore, the NRC-EIL result supported this finding. It shows that users expressed fewer positive emotions such as *joy*, *trust* and *anticipation* but more negative words such as *fear*, *sadness* and *disgust* about weight loss. Notably, *anticipation* is the most frequent positive word expressed in this group, while its paired emotion, *fear*, is the most frequent negative word. This indicates that users are looking forward to losing weight rather than fearing failure on their weight loss journey. Moreover, a time series analysis provided further insight into the sentiment dynamic with time varies on weight loss discussions. The result shows that both positive and negative posts increased, especially in the early months of the pandemic outbreak, from March 2020 to May 2020. However, the dominant sentiment still remained positive. At the same time, all eight emotions show some fluctuations in weight loss behavior. For instance, positive emotions like *joy* and *trust* showed declined trend while negative emotions like *sadness* and *fear* rose sharply at the onset of the pandemic. This indicates the complexity of emotion dynamic expressed on this topic during the pandemic.

Furthermore, sentiment and emotion expressed within specific topics had significant changes during the pandemic. This could suggest that netizens' attitudes and behaviors regarding weight loss changed during the pandemic. While there was a general trend of increasing negative sentiments and a decline in positive sentiments, these changes varied across different weight loss topics. In particular, we found that *Daily updates* (T1), *Diet control* (T14), *Motivation*(T4) and *Emotional eating* (T10) had more positive posts, while *Seeking advice* (T8), *Negative feelings* (T13), *Clothe fit* (T0) and *Fitness App* (T5) had more negative posts during this period. However, the statistical analysis revealed significant sentiment changes across two periods. It is not surprising that most topics showed notable changes, except for topics such as *Weight change* (T2), *Exercise routine* (T6) and *Body image*(T7). For instance, topics such as *Weight loss goal* (T11), *Emotional eating* (T10) and *Motivation* (T4) showed a significant increase in positive posts and a decrease in negative posts. This indicates that users sought effective ways to maintain their motivation or manage emotional challenges related to weight loss. In contrast, topics such as *Calorie tracker* (T14), *Negative feelings* (T13) and *Clothe fit* (T0) experienced a significant decrease in positive posts and an increase in negative posts. This could reflect people could not manage their weight loss effectively due to growing frustrations and challenges caused by the pandemic.

Limitation and future study

Although this study provides findings above through topic modeling and sentiment analysis, several limitations should be acknowledged and could be addressed in future research on weight loss behavior. First, due to the Reddit data restrictions, this study primarily analyzed the historical dataset excluding deleted or removed posts. This selection could miss a large volume of meaningful discussion on weight loss and introduce potential biases into the analysis. Future research, researchers could collect the dataset over a longer duration covering periods before, during, and after the pandemic. Second, this analysis was limited to English discourse only. This means that the finding mainly represents English-speaking users and may lack generalizability to a broader population and cultural background. Future study could compare perceptions and emotions across the different cultural backgrounds during the pandemic. Third, regarding methodology, more advanced machine learning approaches could be used to better track and understand the thematic and sentiment evolution on weight loss such as Structured Topic Models (STM) and Dynamic Topic Models (DTM). Forth, this study only focused on the discussions and sentiments on a specific weight loss subreddit on Reddit. Future research could extend the scope across different subreddits or social media platforms like Twitter or Facebook. And this study is only exploration research to weight loss behavior discussions on social media and the impact of the pandemic on weight loss behavior has not been tested yet. Finally, this study only focuses on textual data, however, social media contains very rich images and videos data that could be further expressed perceptions and emotions. Future research could explore these forms of media to gain a more comprehensive understanding of weight loss discussions.

6.Conclusion

This study has conducted exploration research on weight loss discussion on Reddit from March 2019 to March 2021, particularly in the *loseit* subreddit. By applying the LDA topic modeling and sentiment analysis approach, we observed significant changes in the thematic and sentiment dynamics during the pandemic. The findings highlight changes in public perception regarding weight loss discussions on social media. In this study, we first identified five main areas representing different aspects of weight loss, including Weight Management, Diet, Physical Exercise, Emotions and support, and Appearance. During the pandemic, we found that people discussed more about weight loss (e.g., weight change, calorie tracker, and emotional eating) and upstream topics (e.g., motivation and workout plan). On the other hand, downstream topics (e.g., negative feelings, body image, and clothe fit) of weight loss also evolved. Users talked more about their body image and less about their negative feelings and clothe fit. Sentiment analysis further reveals significant shifts in emotional expression. More positive sentiments were observed than negative ones expressed in weight loss discussions throughout the study. In this weight loss subreddit, users frequently expressed positive emotions like *anticipation* and negative emotions like *fear*. However, the sentiments and emotions varied significantly among most topics. Although this study observed the significant changes in weight loss discussions before and during the pandemic, it does not conclusively determine the impact of the pandemic on these shifts. Future research could consider expanding a broader dataset with multiple social media platforms to capture the impact of this global crisis on public opinions about weight loss.

Appendix – Data

loseit_submissions.zst : The historical dataset contain original post data of 'r/loseit'.

loseit_submissions_2019_2021_ids.txt: The IDs files from 2019 to 2021 exported from the historical dataset.

Submissions.csv: This file contains the post content and its metadata collecting from Reddit.

*Cleaned submissions.csv:*This file contains submission dataset that has been initially cleaned.

*Processed_lda.csv:*The file contains Cleaned submissions that has been text cleaning and preprocessing for LDA algorithm.

*Processed_vader.csv:*The file contains Cleaned submissions that has been text cleaning and preprocessing for VADER analysis.

*Processed_nrc.csv:*The file contains Cleaned submissions that has been text cleaning and preprocessing for NRC-EIL analysis.

*lda_submissions.csv:*This file contains the results of LDA analysis of each post and related metadata.

vader_submissions.csv: This file contains the result of VADER sentiment analysis and weighted sentiment score of each post with related metadata.

nrc_submissions.csv: This file contains the result after applying the NRC-EIL analysis (eight emotions and corresponding scores) of each post with related metadata.

Appendix - Tables

	Pre-pandemic			Dur-pandemic		
Topic	***Positive***	***Negative***	***Netural***	***Positive***	***Negative***	***Netural***
1.Weight Management						
T2	55.45%	42.72%	1.84%	55.17%	43.32%	1.50%
T11	85.41%	13.66%	0.93%	86.77%	12.81%	0.42%
T9	54.07%	44.30%	1.63%	55.97%	42.80%	1.23%
T1	91.33%	7.95%	0.72%	96.75%	3.02%	0.23%
2.Diet						
T10	65.21%	33.40%	1.39%	67.49%	31.36%	1.15%
T12	78.66%	20.32%	1.02%	77.26%	21.92%	0.82%
T3	62.23%	35.49%	2.28%	61.13%	37.55%	1.32%
T14	65.61%	33.40%	0.99%	69.91%	28.85%	1.24%
3.Physical Exercise						
T4	58.12%	39.85%	2.03%	62.11%	36.22%	1.67%
T15	80.75%	18.32%	0.93%	79.51%	19.78%	0.71%
T5	83.77%	15.23%	1.00%	81.07%	18.07%	0.86%
T6	77.87%	21.09%	1.04%	78.26%	20.29%	1.45%
4.Emotions and Support						
T13	68.52%	30.80%	0.67%	65.52%	33.27%	1.21%
T8	80.07%	19.10%	0.83%	76.96%	22.00%	1.04%
5.Appearance						
T7	64.17%	34.14%	1.69%	63.51%	34.33%	2.15%
T0	83.98%	14.48%	1.54%	81.33%	18.04%	0.63%

Table:Proportions of sentiments polarity (positive, negative, netual) by topics across two time periods

Topic	Positive				Negative			
	Joy	*Trust*	*Anticipate*	*Surprise*	*Fear*	*Sadness*	*Disgust*	*Anger*
1.Weight Management								
T2	11.31%	12.64%	16.84%	11.29%	13.78%	13.51%	11.35%	9.28%
T11	14.29%	14.75%	20.40%	9.89%	12.38%	11.38%	8.92%	8.00%
T9	11.65%	14.52%	16.85%	9.27%	14.61%	14.24%	10.12%	8.73%
T1	17.44%	17.26%	19.80%	10.76%	9.98%	9.99%	8.60%	6.16%
2.Diet								
T10	14.14%	16.11%	17.37%	8.93%	12.77%	12.30%	9.90%	8.48%
T12	14.92%	17.24%	18.23%	9.40%	11.90%	11.16%	9.15%	8.00%
T3	10.62%	16.16%	16.89%	10.46%	12.80%	13.68%	11.40%	7.99%
T14	12.76%	15.26%	18.08%	9.74%	13.17%	12.61%	9.99%	8.39%
3.Physical Exercise								
T4	12.19%	14.33%	18.59%	9.11%	12.83%	13.17%	10.38%	9.40%
T15	13.18%	16.31%	19.75%	9.46%	12.28%	11.93%	9.45%	7.64%
T5	12.83%	18.39%	18.96%	9.52%	11.96%	12.10%	9.40%	6.84%
T6	12.41%	17.36%	20.68%	9.02%	12.16%	11.82%	8.04%	8.51%
4.Emotions and Support								
T13	13.46%	15.45%	16.75%	9.29%	13.44%	13.02%	9.81%	8.78%
T8	12.16%	15.46%	16.75%	10.60%	13.71%	13.29%	10.15%	7.88%
5.Appearance								
T7	11.85%	13.38%	14.48%	10.17%	13.62%	14.74%	13.19%	8.57%
T0	13.25%	15.77%	18.24%	10.28%	12.58%	11.73%	9.69%	8.46%

Table:Proportions of eight emotion distribution by topics before the pandemic

Topic	**Positive**				**Negative**			
	Joy	***Trust***	***Anticipate***	***Surprise***	***Fear***	***Sadness***	***Disgust***	***Anger***
1.Weight Management								
T2	11.20%	12.69%	16.81%	11.30%	14.00%	13.54%	11.35%	9.10%
T11	14.21%	15.07%	19.77%	9.92%	12.87%	11.46%	8.91%	7.79%
T9	11.75%	14.29%	16.85%	9.47%	14.83%	14.18%	10.27%	8.38%
T1	20.40%	19.59%	21.86%	10.34%	8.89%	7.83%	6.75%	4.34%
2.Diet								
T10	14.31%	16.11%	17.69%	8.98%	12.82%	12.25%	9.75%	8.09%
T12	14.89%	17.01%	18.17%	9.44%	12.09%	11.33%	9.32%	7.75%
T3	10.61%	16.12%	16.81%	10.31%	13.23%	13.70%	11.31%	7.90%
T14	13.10%	15.30%	18.43%	9.76%	13.07%	12.29%	10.17%	7.88%
3.Physical Exercise								
T4	12.05%	13.72%	18.49%	9.29%	13.44%	13.37%	10.21%	9.43%
T15	12.74%	15.99%	19.53%	9.36%	12.85%	12.18%	9.43%	7.92%
T5	12.70%	17.85%	18.38%	9.48%	12.54%	12.39%	9.63%	7.04%
T6	12.94%	16.52%	20.74%	8.88%	12.35%	12.07%	8.16%	8.36%
4.Emotions and Support								
T13	13.12%	15.40%	16.60%	9.10%	13.85%	13.31%	9.82%	8.79%
T8	12.29%	15.39%	17.01%	10.44%	13.90%	13.24%	9.91%	7.83%
5.Appearance								
T7	11.76%	13.18%	14.41%	10.18%	13.91%	14.94%	13.34%	8.30%
T0	13.07%	15.48%	18.23%	10.24%	12.94%	11.94%	9.72%	8.38%

Table:Proportions of eight emotion distribution by topics during the pandemic

Appendix - Code

The code was written and run in Jupyter Notebook 7.0.8 (Kluyver, et al., 2016) under Python 3.9.12 kernel (Van Rossum & Drake, 2009).

1.Data Collection

a. Extract loseit submissions IDs from Reddit historical data.

```
# This section of the code is adapted from an external script available at:
# https://github.com/Watchful1/PushshiftDumps/blob/master/scripts/filter_file.py

# 1.Define file paths and data extraction parameters

# path to the input/output file
input_file = r"/Users/arptns/Desktop/Reddit data/loseit_submissions.zst"
output_file = r"/Users/arptns/Desktop/Reddit data/loseit_submissions_2019_2021_ids"

# date range for filtering the data
from_date = datetime.strptime("2019-03-01", "%Y-%m-%d") # start date
to_date = datetime.strptime("2021-04-01", "%Y-%m-%d") # end date

# specify the output file format
output_format = 'txt'

# filter data by subreddit
field, values = "subreddit", ["loseit"]

# define the data field to be extracted
single_field = 'id'

# enable logging of errors during data processing
write_bad_lines = True

# File containing additional filter values
values_file = None

# control matching strictness in data filtering
exact_match = False

# 2.Logging setup to capture errors during processing

log = logging.getLogger("bot")
log.setLevel(logging.INFO)
log_formatter = logging.Formatter('%(asctime)s - %(levelname)s: %(message)s')
log_str_handler = logging.StreamHandler()
log_str_handler.setFormatter(log_formatter)
log.addHandler(log_str_handler)
if not os.path.exists("logs"):
    os.makedirs("logs")
log_file_handler = logging.handlers.RotatingFileHandler(os.path.join("logs", "bot.log"), maxBytes=1024*1024*16,
backupCount=5)
log_file_handler.setFormatter(log_formatter)
log.addHandler(log_file_handler)
```

3.Function definitions to handle data extraction and processing

```
def write_line_zst(handle, line):
    handle.write(line.encode('utf-8'))
    handle.write("\n".encode('utf-8'))


def write_line_json(handle, obj):
    handle.write(json.dumps(obj))
    handle.write("\n")


def write_line_single(handle, obj, field):
    if field in obj:
        handle.write(obj[field])
    else:
        log.info(f"{field} not in object {obj['id']}")
    handle.write("\n")


def write_line_csv(writer, obj, is_submission):
    output_list = []
    output_list.append(str(obj['score']))
    output_list.append(datetime.fromtimestamp(int(obj['created_utc'])).strftime("%Y-%m-%d"))
    if is_submission:
        output_list.append(obj['title'])
    output_list.append(f"u/{obj['author']}")
    output_list.append(f"https://www.reddit.com{obj['permalink']}")
    if is_submission:
        if obj['is_self']:
            if 'selftext' in obj:
                output_list.append(obj['selftext'])
            else:
                output_list.append("")
        else:
            output_list.append(obj['url'])
    else:
        output_list.append(obj['body'])
    writer.writerow(output_list)

def read_and_decode(reader, chunk_size, max_window_size, previous_chunk=None, bytes_read=0):
    chunk = reader.read(chunk_size)
    bytes_read += chunk_size
    if previous_chunk is not None:
        chunk = previous_chunk + chunk
    try:
        return chunk.decode()
    except UnicodeDecodeError:
        if bytes_read > max_window_size:
            raise UnicodeError(f"Unable to decode frame after reading {bytes_read:,} bytes")
        log.info(f"Decoding error with {bytes_read:,} bytes, reading another chunk")
        return read_and_decode(reader, chunk_size, max_window_size, chunk, bytes_read)


def read_lines_zst(file_name):
    with open(file_name, 'rb') as file_handle:
        buffer = ''
        reader = zstandard.ZstdDecompressor(max_window_size=2**31).stream_reader(file_handle)
        while True:
            chunk = read_and_decode(reader, 2**27, (2**29) * 2)

            if not chunk:
```

```
                break
            lines = (buffer + chunk).split("\n")

            for line in lines[:-1]:
                yield line.strip(), file_handle.tell()

            buffer = lines[-1]

        reader.close()


def process_file(input_file, output_file, output_format, field, values, from_date, to_date, single_field, exact_match):
    output_path = f"{output_file}.{output_format}"
    is_submission = "submission" in input_file
    log.info(f"Input: {input_file} : Output: {output_path} : Is submission {is_submission}")
    writer = None
    if output_format == "zst":
        handle = zstandard.ZstdCompressor().stream_writer(open(output_path, 'wb'))
    elif output_format == "txt":
        handle = open(output_path, 'w', encoding='UTF-8')
    elif output_format == "csv":
        handle = open(output_path, 'w', encoding='UTF-8', newline='')
        writer = csv.writer(handle)
    else:
        log.error(f"Unsupported output format {output_format}")
        sys.exit()

    file_size = os.stat(input_file).st_size
    created = None
    matched_lines = 0
    bad_lines = 0
    total_lines = 0
    for line, file_bytes_processed in read_lines_zst(input_file):
        total_lines += 1
        if total_lines % 100000 == 0:
            log.info(f"{created.strftime('%Y-%m-%d %H:%M:%S')} : {total_lines:,} : {matched_lines:,} : {bad_lines:,} :
{file_bytes_processed:,}:{(file_bytes_processed / file_size) * 100:.0f}%")

        try:
            obj = json.loads(line)
            created = datetime.utcfromtimestamp(int(obj['created_utc']))

            if created < from_date:
                continue
            if created > to_date:
                continue

            if field is not None:
                field_value = obj[field].lower()
                matched = False
                for value in values:
                    if exact_match:
                        if value == field_value:
                            matched = True
                            break
                    else:
                        if value in field_value:
                            matched = True
                            break
                if not matched:
                    continue
```

```python
            matched_lines += 1
            if output_format == "zst":
                write_line_zst(handle, line)
            elif output_format == "csv":
                write_line_csv(writer, obj, is_submission)
            elif output_format == "txt":
                if single_field is not None:
                    write_line_single(handle, obj, single_field)
                else:
                    write_line_json(handle, obj)
            else:
                log.info(f"Something went wrong, invalid output format {output_format}")
        except (KeyError, json.JSONDecodeError) as err:
            bad_lines += 1
            if write_bad_lines:
                if isinstance(err, KeyError):
                    log.warning(f"Key {field} is not in the object: {err}")
                elif isinstance(err, json.JSONDecodeError):
                    log.warning(f"Line decoding failed: {err}")
                log.warning(line)

    handle.close()
    log.info(f"Complete : {total_lines:,} : {matched_lines:,} : {bad_lines:,}")

# Main function to orchestrate data processing based on user-defined settings

if __name__ == "__main__":
    if single_field is not None:
        log.info("Single field output mode, changing output file format to txt")
        output_format = "txt"

    if values_file is not None:
        values = []
        with open(values_file, 'r') as values_handle:
            for value in values_handle:
                values.append(value.strip().lower())
        log.info(f"Loaded {len(values)} from values file {values_file}")
    else:
        values = [value.lower() for value in values]  # convert to lowercase

    log.info(f"Filtering field: {field}")
    if len(values) <= 20:
        log.info(f"On values: {','.join(values)}")
    else:
        log.info(f"On values:")
        for value in values:
            log.info(value)
    log.info(f"Exact match {('on' if exact_match else 'off')}. Single field {single_field}.")
    log.info(f"From date {from_date.strftime('%Y-%m-%d')} to date {to_date.strftime('%Y-%m-%d')}")
    log.info(f"Output format set to {output_format}")

    input_files = []
    if os.path.isdir(input_file):
        if not os.path.exists(output_file):
            os.makedirs(output_file)
        for file in os.listdir(input_file):
            if not os.path.isdir(file) and file.endswith(".zst"):
                input_name = os.path.splitext(os.path.splitext(os.path.basename(file))[0])[0]
                input_files.append((os.path.join(input_file, file), os.path.join(output_file, input_name)))
    else:
        input_files.append((input_file, output_file))
    log.info(f"Processing {len(input_files)} files")
```

```
    for file_in, file_out in input_files:
        process_file(file_in, file_out, output_format, field, values, from_date, to_date, single_field, exact_match)
```

b. Scrape posts and metadata from Reddit.

```
# Import the 'submission ids' file
submission_ids = pd.read_csv('/Users/arptns/Desktop/Reddit data/loseit_submissions_2019_2021_ids.txt', header=None,
names=['sub_id'])

len(submission_ids) # 115123 ids

# Set the Reddit API access with PRAW
reddit = praw.Reddit(client_id='F2PirR2UnatbMaFiVlJlIg',
                     client_secret='hcMIMgNKI4vvatkV4Qsc2SjLuDKyhA',
                     user_agent='web:Joeeyy2046:v1.0 (by /u/Joeeyy2046)')

subreddit = reddit.subreddit('loseit') # set the subreddit
batch_size = 10000 # set 10000 submissions per batch
submissions = [] # create an empty list to store submissions dataset

# For loop to scrape the submissions through Reddit API

for i in range(0, len(submission_ids), batch_size):
    batch_ids = submission_ids[i:i+batch_size]  # retrieve current batch of submissions IDs
    for sub_id in batch_ids['sub_id']:
        try:
            # fetch the submissions data from Reddit
            submission = reddit.submission(id=sub_id)
            submission.comments.replace_more(limit=0) # remove comments beyond the view limit
            created_time = datetime.utcfromtimestamp(submission.created_utc).strftime('%Y-%m-%d %H:%M:%S')

            # append the submission details to the list
            submissions.append({
                'Creation Time': created_time,
                'Title': submission.title,
                'User': str(submission.author),
                'Selftext': submission.selftext,
                'Score': submission.score,
                'Number of Comments': submission.num_comments,
                'URL': submission.url
            })
        except Exception as e:
            print(f"Error processing post {sub_id}: {e}")

        time.sleep(1) # avoid exceeding the rate limits

    # convert the current batch of submissions into a dataFrame and save to CSV file
    batch_df = pd.DataFrame(submissions)
    batch_df.to_csv(f'/Users/arptns/Desktop/Reddit data/submissions_2019_2021/submissions_batch_{i}.csv', index=False)

# after the process finished, merge and save into a single dataframe
submissions_df = pd.DataFrame(submissions)
submissions_df.to_csv('/Users/arptns/Desktop/Reddit data/submissions.csv', index=False)

print("All the submissions are finished.")
```

c. Filter out any removed or deleted posts.

```
# Open the 'r/loseit' submissions dataset
submissions = pd.read_csv('/Users/arptns/Desktop/Reddit data/submissions.csv')

# Count the deleted submissions
```

```
submissions['Selftext'].str.contains('\[deleted\]', na=False).sum() # 21515 submissions are deleted

# Count the removed submissions
submissions['Selftext'].str.contains('\[removed\]', na=False).sum() # 52755 submissions are removed

# Replace the deleted and removed submissions to NA and remove them
submissions['Selftext'].replace(['[deleted]', '[removed]'], np.nan, inplace=True)
submissions.dropna(subset=['Selftext'], inplace=True)

# Remove posts from automoderator
submissions = submissions[submissions['Author'] != 'AutoModerator']

# Remove posts from official challenges
submissions = submissions[~submissions['Title'].str.contains(r'\[Challenge\]', na=False)]

# Count the total number of submissions
submissions['Selftext'].count() # 35384 submissions

# Reset the index
submissions = submissions.reset_index(drop=True)

cleaned_submissions = submissions.copy()

# Convert 'Creation Time' column to datetime type
cleaned_submissions['Creation Time'] = pd.to_datetime(cleaned_submissions['Creation Time'])

# Add new column 'Period' to pre/dur pandemic
cleaned_submissions['Period'] = np.where(
    cleaned_submissions['Creation Time'] < pd.Timestamp('2020-03-11'),
    'pre_pandemic',
    'dur_pandemic'
)

# Save the cleaned submissions to CSV file
cleaned_submissions.to_csv('/Users/arptns/Desktop/Reddit data/cleaned_submissions.csv', index=False)
```

d. Basic statistics overview and visualization.

```
cleaned_submissions = pd.read_csv('/Users/arptns/Desktop/Reddit data/cleaned_submissions.csv')

pre_pandemic = cleaned_submissions[cleaned_submissions['Period'] == 'pre_pandemic']
len(cleaned_submissions[cleaned_submissions['Period'] == 'pre_pandemic']) # 16244 posts before the pandemic

dur_pandemic = cleaned_submissions[cleaned_submissions['Period'] == 'dur_pandemic']
len(cleaned_submissions[cleaned_submissions['Period'] == 'dur_pandemic']) # 19140 posts after the pandemic

# Basic Statistics overview of the loseit dataset before the pandemic

# Mean
pre_pandemic_mean_posts = pre_pandemic['Creation Time'].dt.date.value_counts().mean()
pre_pandemic_mean_scores = pre_pandemic['Score'].mean()
pre_pandemic.loc[:, 'Words per post'] = pre_pandemic['Selftext'].str.split().apply(len)
pre_pandemic_mean_words = pre_pandemic['Words per post'].mean()

# Median
pre_pandemic_median_posts = pre_pandemic['Creation Time'].dt.date.value_counts().median()
pre_pandemic_median_scores = pre_pandemic['Score'].median()
pre_pandemic_median_words = pre_pandemic['Words per post'].median()

# SD
pre_pandemic_sd_posts = pre_pandemic['Creation Time'].dt.date.value_counts().std()
pre_pandemic_sd_scores = pre_pandemic['Score'].std()
```

```
pre_pandemic_sd_words = pre_pandemic['Words per post'].std()

# Print the output
print("Pre-pandemic era:")
print(f"Posts per day - Mean: {pre_pandemic_mean_posts}, Median: {pre_pandemic_median_posts}, SD:
{pre_pandemic_sd_posts}")
print(f"Scores per post - Mean: {pre_pandemic_mean_scores}, Median: {pre_pandemic_median_scores}, SD:
{pre_pandemic_sd_scores}")
print(f"Words per post - Mean: {pre_pandemic_mean_words}, Median: {pre_pandemic_median_words}, SD:
{pre_pandemic_sd_words}")

# Basic Statistics overview of the loseit dataset during the pandemic

# Mean
dur_pandemic_mean_posts = dur_pandemic['Creation Time'].dt.date.value_counts().mean()
dur_pandemic_mean_scores = dur_pandemic['Score'].mean()
dur_pandemic.loc[:, 'Words per post'] = dur_pandemic['Selftext'].str.split().apply(len)
dur_pandemic_mean_words = dur_pandemic['Words per post'].mean()

# Median
dur_pandemic_median_posts = dur_pandemic['Creation Time'].dt.date.value_counts().median()
dur_pandemic_median_scores = dur_pandemic['Score'].median()
dur_pandemic_median_words = dur_pandemic['Words per post'].median()

# SD
dur_pandemic_sd_posts = dur_pandemic['Creation Time'].dt.date.value_counts().std()
dur_pandemic_sd_scores = dur_pandemic['Score'].std()
dur_pandemic_sd_words = dur_pandemic['Words per post'].std()

# Plot the submissions distribution monthly

# Convert 'Creation Time' column to datetime type
cleaned_submissions['Creation Time'] = pd.to_datetime(cleaned_submissions['Creation Time'],
format='%Y-%m-%d %H:%M:%S') # specify datetime format

# Extract 'Year', 'Month', 'Day' information
cleaned_submissions['Year'] = cleaned_submissions['Creation Time'].dt.year
cleaned_submissions['Month'] = cleaned_submissions['Creation Time'].dt.month
cleaned_submissions['Day'] = cleaned_submissions['Creation Time'].dt.day

# Count monthly submissions
subs_per_month = cleaned_submissions.groupby(['Year', 'Month']).size().reset_index(name='Post Count')

# Plot the monthly submissions
plt.figure(figsize=(15, 6))
plt.bar(subs_per_month['Year'].astype(str) + '-' + subs_per_month['Month'].astype(str), subs_per_month['Post Count'],
color='#1f77b4')
plt.ylabel('Number of Posts', fontsize=13)
plt.xticks(rotation=45, fontsize=13)
plt.yticks(rotation=90, fontsize=13)
plt.tight_layout()
plt.show()
```

2.Data preprocess

a. Clean the submission’s dataset.

```
# Define VADER cleaning function
def clean_vader(text):
    text = re.sub(r'https?://\S+|www\.\S+', '', text)
```

```
    text = re.sub(r'&#x200B;|\bx\s*b\b', '', text)
    text = re.sub(r'\n', ' ', text)
    return text

# Define LDA and NRC-EIL cleaning function
def clean_lda_nrceil(text):
    text = re.sub(r'https?://\S+|www\.\S+', '', text)
    text = re.sub(r'&#x200B;|\bx\s*b\b', '', text)
    text = re.sub(r'\bdon t\b'', ''do not', '', text)
    text = re.sub(r'\bdidn t\b'', ''did not', '', text)
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    text = text.lower()
    text = re.sub(r'\n', ' ', text)
    return text

# Apply cleaning functions
cleaned_submissions['clean_vader'] = cleaned_submissions['Selftext'].apply(clean_vader)
cleaned_submissions['clean_lda_nrceil'] = cleaned_submissions['Selftext'].apply(clean_lda_nrceil)
```

b. Preprocess the submission’s dataset.

```
## Topic modeling (LDA)

# Step 1: Tokenize the words
cleaned_submissions['tokenized_lda'] = [
    word_tokenize(text) for text in cleaned_submissions['clean_lda_nrceil']
]

# Step 2: Lemmatization

nlp = spacy.load('en_core_web_sm', disable=['parser', 'ner'])
def lemmatize_text(tokens, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV', 'PROPN']):
    doc = nlp(' '.join(tokens))
    return [token.lemma_ for token in doc if token.pos_ in allowed_postags]
cleaned_submissions['lemmatized_lda_texts'] = [
    lemmatize_text(words) for words in cleaned_submissions['tokenized_lda']
]

# Step 3: Remove customed and stardard stop words
stop_words = set(stopwords.words('english')) # load stop words
cleaned_submissions['lemmatized_lda_texts'] = [
    [word for word in words if word.lower() not in stop_words] for words in cleaned_submissions['lemmatized_lda_texts']
]

# Step 4: Remove the word with less than 3
cleaned_submissions['preprocessed_lda'] = [
    ' '.join([word for word in words if len(word) >= 3]) for words in cleaned_submissions['lemmatized_lda_texts']]

## Sentiment analysis (VADER)

# Step 1: Tokenize the words
cleaned_submissions['tokenized_vader'] = [
    word_tokenize(text) for text in cleaned_submissions['clean_vader']
]

# Step 2: Remove stop words
stop_words = set(stopwords.words('english')) # load stop words
cleaned_submissions['vader_texts'] = [
    [word for word in tokens if word not in stop_words] for tokens in cleaned_submissions['tokenized_vader']
]

# Step 3: Convert the list of words back into text.
```

```
cleaned_submissions['preprocessed_vader'] = cleaned_submissions['vader_texts'].apply(lambda words: ' '.join(words))

## Sentiment analysis (NRC-EIL)

# Step 1: Tokenize the words
cleaned_submissions['tokenized_nrc'] = [
    word_tokenize(text) for text in cleaned_submissions['clean_lda_nrceil']
]

# Step 2: Lemmatization
def lemmatize_text(tokens):
    doc = nlp(' '.join(tokens))
    return [token.lemma_ for token in doc]
cleaned_submissions['lemmatized_nrc_texts'] = [
    lemmatize_text(words) for words in cleaned_submissions['tokenized_nrc']
]

# Step 2: Remove stop words
stop_words = set(stopwords.words('english')) # load stop words
cleaned_submissions['lemmatized_nrc_texts'] = [
    [word for word in words if word.lower() not in stop_words] for words in cleaned_submissions['lemmatized_nrc_texts']
]

# Step 3: Convert the list of words back into text
cleaned_submissions['preprocessed_nrc'] = cleaned_submissions['lemmatized_nrc_texts'].apply(lambda words: ' '.join(words))
```

c. Save the preprocessed submissions for analysis.

```
# Convert the preprocessed submissions for LDA to data frame
cleaned_submissions['Selftext'] = cleaned_submissions['preprocessed_lda'] preprocessed_lda = cleaned_submissions[['Creation Time', 'Selftext', 'Score', 'Number of Comments', 'Period']].copy()

# Convert the preprocessed submissions for VADER to data frame
cleaned_submissions['Selftext'] = cleaned_submissions['preprocessed_vader']
preprocesseds_vader = cleaned_submissions[['Creation Time', 'Selftext', 'Score', 'Number of Comments', 'Period']].copy()

# Convert the preprocessed submissions for NRC-EIL to data frame
cleaned_submissions['Selftext'] = cleaned_submissions['preprocessed_nrc']
preprocessed_nrc = cleaned_submissions[['Creation Time', 'Selftext', 'Score', 'Number of Comments', 'Period']].copy()

# Find preprocess value with any NaN values
na_lda = preprocessed_lda[preprocessed_lda.isna().any(axis=1)]
na_vader = preprocessed_vader[preprocessed_vader.isna().any(axis=1)]
na_nrc = preprocessed_nrc[preprocessed_nrc.isna().any(axis=1)]
# posts with 5696,13334,31352 ids has NAN values in LDA and NRC preprocessing data

# Remove this three posts with NaN in the 'Selftext' column
indices_to_drop = [5696, 13334, 31352] # 35381 submissions left in the datasets
preprocessed_lda = preprocessed_lda.drop(indices_to_drop)
preprocessed_vader = preprocessed_vader.drop(indices_to_drop)
preprocessed_nrc = preprocessed_nrc.drop(indices_to_drop)

# Reset the index
preprocessed_lda = preprocessed_lda.reset_index(drop=True)
preprocessed_vader = preprocessed_vader.reset_index(drop=True)
preprocessed_nrc = preprocessed_nrc.reset_index(drop=True)

# Save the preprocessed dataset
preprocessed_lda.to_csv('/Users/arptns/Desktop/Reddit data/preprocessed_lda.csv', index=False)
preprocessed_vader.to_csv('/Users/arptns/Desktop/Reddit data/preprocessed_vader.csv', index=False)
preprocessed_nrc.to_csv('/Users/arptns/Desktop/Reddit data/preprocessed_nrc.csv', index=False)
```

3.Topic Modeling

a. Build dictionary and BOW corpus for LDA model.

```
# Open the dataset for LDA analysis
preprocessed_lda = pd.read_csv('/Users/arptns/Desktop/Reddit data/preprocessed_lda.csv')

# Covert the post text to string
lda_texts = preprocessed_lda['Selftext'].astype(str)
tokenized_texts = [[word for word in word_tokenize(doc) if word.lower() not in stop_words] for doc in lda_texts]

# Create the dictionary
dictionary = corpora.Dictionary(tokenized_texts)
dictionary.filter_extremes(no_below=5, no_above=0.8) # Filter extreme words in the dictionary
print(dictionary)

# Create the corpus
corpus = [dictionary.doc2bow(doc) for doc in tokenized_texts]
```

b. Test the optimal topics numbers from 5 to 30.

```
# Test the number of different topics
topic_numbers = range(5, 31, 1)

# Create an empty list of coherence scores
c_v_scores = []

# Configure LDA multicore model
alpha = 1  # Hyperparameter for document-topic distribution
beta = 0.1  # Hyperparameter for topic-word distribution
random_state = 42  # Random state for reproducibility
workers = 3  # Number of CPUs to use
passes = 20  # Number of passes through the corpus
iterations = 1000 # interation 1000 times

# Test the optimal topics
for num_topics in topic_numbers:
    lda_model = LdaMulticore(corpus=corpus, id2word=dictionary, num_topics=num_topics, alpha=alpha, eta=beta,
workers=workers, passes=passes, random_state=random_state, iterations=iterations)
    c_v_model = CoherenceModel(model=lda_model, texts=tokenized_texts, dictionary=dictionary, coherence='c_v')
    c_v_lda = c_v_model.get_coherence()
    c_v_scores.append(c_v_lda)

# Consistency scores visualization
topic_numbers = range(5, 31) # convert the range object to a list
# Plot Pre-pandemic consistency scores
plt.plot(topic_numbers, c_v_scores, label='Coherence Score')
plt.axvline(x=16, color='red', linestyle='--', label='Selected Topic Number')
plt.legend(loc='lower right', prop={'size': 8}) # Add legend
plt.xticks([5, 10, 15, 20, 25, 30])  # Set the ticks of x-axis
plt.xlabel('Number of Topics')
plt.ylabel('Coherence Score')
plt.grid(True, color='gray', linestyle='-', linewidth=0.5)
plt.show()

# Print the coherence_score with number of topics
for num_topics, coherence_score in zip(topic_numbers, c_v_scores):
    print(f'The model with num_topics = {num_topics} has a coherence value of {coherence_score:.3f}')
```

c. Run the LDA model with the optimal topic numbers.

```
# Step 1.Build new dictionary and BOW corpus

# add Custom stop words
custom_stop_words = set(['get','make','take','really','still','even','also','try','see','thing', 'new', 'end', 'much'])

# tokenize new text
tokenized_texts_16 = [[word for word in word_tokenize(doc) if word.lower() not in custom_stop_words] for doc in lda_texts]

# create new dictionary
dictionary_16 = corpora.Dictionary(tokenized_texts_16)
dictionary_16.filter_extremes(no_below=5, no_above=0.8)
corpus_16 = [dictionary_16.doc2bow(doc) for doc in tokenized_texts_16]

# Step 2.Run the LDA model

lda_model_16 = LdaMulticore(corpus=corpus_16,
                    id2word=dictionary_16,
                    num_topics=16,
                    alpha=alpha,
                    eta=beta,
                    random_state=random_state,
                    workers=workers,
                    iterations = iterations,
                    passes=passes)

# Print each topic number and its associated words
for idx, topic in lda_model_16.print_topics(-1):
    print('Topic: {} \nWords: {}'.format(idx, topic))

# Calculate new coherence score for the LDA model
coherence_lda_model_16 = CoherenceModel(model=lda_model_16, texts=tokenized_texts_16, dictionary=dictionary_16,
coherence='c_v')
coherence_lda_16 = coherence_lda_model_16.get_coherence()
print('\nCoherence Score: ', coherence_lda_16)

# Step 3.Save the LDA submissions

# get topic distribution for each document in the corpus
doc_topics = [lda_model_16.get_document_topics(doc) for doc in corpus_16]

# find the topic with the highest probability for each document
relevant_topics = [max(doc, key=lambda x: x[1])[0] for doc in doc_topics] # contains the most relevant topic ID for each
document

# add a new column 'Topic' to the DataFrame
preprocessed_lda['Topic'] = relevant_topics

lda_submissions = preprocessed_lda.copy()

lda_submissions.to_csv('/Users/arptns/Desktop/Reddit data/lda_submissions.csv', index=False)

# Step 4.Visualize the proportions of each LDA topic

# Calculate the proportions of each topic
topic_counts_total = lda_submissions['Topic'].value_counts() # the total number of posts of each topic
topic_proportions_total = topic_counts_total / topic_counts_total.sum() * 100 # the proportions of each topic
topic_proportions_total

topic_labels = ['T' + str(topic) for topic in topic_counts_total.index]
topic_counts_sorted = topic_counts_total.sort_values(ascending=False) # Sort the data
```

```
topic_labels_sorted = ['T' + str(topic) for topic in topic_counts_sorted.index]

# Create the plot
plt.figure(figsize=(20, 8))
sns.barplot(x=topic_labels_sorted, y=topic_counts_sorted.values, color='#1f77b4')
plt.ylabel('Number of Documents', fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=13)
plt.gca().yaxis.set_major_formatter(FuncFormatter(lambda x, _: f'{int(x/1000)}k'))
plt.show()
```

d. Temporal analysis by topics in five themes.

```
# Assuming 'Date' is the name of the column containing datetime information
lda_submissions['Creation Time'] = pd.to_datetime(lda_submissions['Creation Time'])

# Calculate the number of posts in each topic by two periods
topic_counts = lda_submissions.groupby(['Period', 'Topic']).size().unstack(fill_value=0)

# Calculate the proportions of posts in each topic by two periods
total_posts = lda_submissions.shape[0] # total posts for the dataset
topic_counts = topic_counts.stack().reset_index(name='Count') # posts in each topic by two periods
topic_proportions = topic_counts.copy()
topic_proportions['Proportion'] = (topic_proportions['Count'] / total_posts) * 100

# Set the 'Creation Time' column as the index
lda_submissions.set_index('Creation Time', inplace=True)

# Apply the resample method to resample by month and calculate the number of posts
monthly_posts = lda_submissions.resample('M').size()

# Initialize a DataFrame to store the expected proportions for each topic
topic_proportion = pd.DataFrame(index=monthly_posts.index)

# Calculate the expected proportion for each topic
for topic in lda_submissions['Topic'].unique():
    # Filter data for the specific topic and resample by month
    monthly_topic_posts = lda_submissions[lda_submissions['Topic'] == topic].resample('M').size()
    # Calculate the proportion by dividing the number of posts for each topic by the total number of posts for each month
    topic_proportion[topic] = monthly_topic_posts / monthly_posts

# Normalize the topic counts to get proportions
topic_proportion = topic_proportion.div(topic_proportion.sum(axis=1), axis=0)

# Visualize the topic trends over time
def plot_topic_trends(topic_proportion, topics_to_plot, names):
    # Convert the time index (date) to a number
    dates_numeric = np.arange(len(topic_proportion.index))

    # Ensure the topic_proportion index is adjusted to the start of each month
    adjusted = topic_proportion.index - pd.offsets.MonthEnd(1) + pd.Timedelta(days=1)

    # Set the figure size
    plt.figure(figsize=(20, 5))

    # Loop through each topic to plot
    for i, topic in enumerate(topics_to_plot):
        # Get the expected ratio sequence of the current theme
        y = topic_proportion[topic].values * 100
        # Apply LOESS regression
        lowess_results = sm.nonparametric.lowess(y, dates_numeric, frac=0.4) # frac parameter controls the smoothness
        # Draw the original data points
```

```
            plt.scatter(adjusted, y, alpha=0.4, s=8)
            # Draw the smooth curve
            plt.plot(adjusted, lowess_results[:, 1], label=names[i], lw=1.5)

    # Mark the pandemic start point with a vertical line
    plt.axvline(x=adjusted[12], color='red', linestyle='--', label='Pandemic start point')
    plt.ylabel('Topic Proportion', fontsize=18)
    plt.gca().yaxis.set_major_formatter(FuncFormatter(lambda y, _: '{:.0f}%'.format(y))) # format y-axis as percentages

    # Ensure x-axis shows every month
    plt.gca().xaxis.set_major_locator(mdates.MonthLocator())
    plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%Y-%m'))
    plt.tick_params(axis='both', which='both', length=0) # remove the small tick lines on x and y axes
    plt.legend(loc='upper right', fontsize=10) # add legend
    plt.xticks(rotation=45, fontsize=13)  # rotate x-axis labels
    plt.tight_layout()
    plt.grid(True, color='lightgray')  # add grid
    plt.show()

# Plot the weight management theme
topics_to_plot_1 = [2, 11, 9, 1]
names_1 = ['Weight change (T2)', 'Weight loss goal (T11)', 'Medication (T9)', 'Daily updates(T1)']
plot_topic_trends(topic_proportion, topics_to_plot_1, names_1)

# Plot the Diet theme
topics_to_plot_2 = [10, 12, 3, 14]
names_2 = ['Emotional eating (T10)', 'Food choice (T12)', 'Calorie tracker (T3)', 'Diet control(T14)']
plot_topic_trends(topic_proportion, topics_to_plot_2, names_2)

# Plot the Physical Exercise theme
topics_to_plot_3 = [4, 15, 5, 6]
names_3 = ['Motivation (T4)', 'Workout plan (T15)', 'Fitness App (T5)', 'Exercise routine (T6)']
plot_topic_trends(topic_proportion, topics_to_plot_3, names_3)

# Plot the Emotions and support theme
topics_to_plot_4 = [13, 8]
names_4 = ['Negative feelings (T13)', 'Seeking advice (T8)']
plot_topic_trends(topic_proportion, topics_to_plot_4)

# Plot the Appearance theme
topics_to_plot_5 = [7, 0]
names_5 = ['Body image (T7)', 'Clothe fit (T0)']
plot_topic_trends(topic_proportion, topics_to_plot_5)
```

4.Sentiment Analysis (VADER)

<u>a. Analyze Vader sentiment polarity on each post.</u>

```
# Instantiate the VADER analyzer
sid = SentimentIntensityAnalyzer()

# Analyze sentiment on each post
def analyze_vader_sentiment(text):
    text = str(text) # convert text to string type
    scores = sid.polarity_scores(text) # get sentiment score
    # determine the emotion category based on the comprehensive score
    if scores['compound'] > 0.15:
        sentiment = 'positive' # if compound scores > 0.15 then positive sentiment
    elif scores['compound'] < -0.15: # if compound scores < -0.14 then negative sentiment
        sentiment = 'negative'
```

```
    elif scores['compound'] >= -0.15 and scores['compound'] <= 0.15:
        sentiment = 'neutral' # otherwise netural sentiment for -0.15< compound < 0.15
    # return the sentiment category and score
    return sentiment, scores['pos'], scores['neg'], scores['neu'], scores['compound']
```

b.VADER descriptive statistics.

```
# Count the number of negative, positive, and neutral posts in each time period
sentiment_counts = vader_submissions.groupby(['Period', 'Sentiment']).size()#.unstack(fill_value=0)

# Filter the data for pre-pandemic and during-pandemic periods
pre_pandemic_posts = vader_submissions[vader_submissions['Period'] == 'pre_pandemic']
dur_pandemic_posts = vader_submissions[vader_submissions['Period'] == 'dur_pandemic']

# Count the number of positive, negative, and neutral posts in each period
sentiment_counts_pre = pre_pandemic_posts['Sentiment'].value_counts()
sentiment_counts_dur = dur_pandemic_posts['Sentiment'].value_counts()

# Calculate the percentage of each sentiment category within each period
sentiment_percentages_pre = sentiment_counts_pre / sentiment_counts_pre.sum() * 100
sentiment_percentages_dur = sentiment_counts_dur / sentiment_counts_dur.sum() * 100
```

c.Visualize sentiment changes over time.

```
# Convert 'Creation Time' to datetime
vader_submissions['Creation Time'] = pd.to_datetime(vader_submissions['Creation Time'])
# Convert 'Creation Time' to month
vader_submissions['Month'] = vader_submissions['Creation Time'].dt.to_period('M')
# Caculate the average number of posts monthly
monthly_avg_pov = vader_submissions[vader_submissions['Sentiment'] == 'positive'].groupby('Month').size()
monthly_avg_neg = vader_submissions[vader_submissions['Sentiment'] == 'negative'].groupby('Month').size()
monthly_avg_neu = vader_submissions[vader_submissions['Sentiment'] == 'neutral'].groupby('Month').size()

# Plot the figures
plt.figure(figsize=(20, 5))
plt.plot(monthly_avg_pov.index.astype(str), monthly_avg_pov.values, marker='o', color='#ff7f0e', label='Positive')
plt.plot(monthly_avg_neg.index.astype(str), monthly_avg_neg.values, marker='x', color='#1f77b4', label='Negative')
plt.plot(monthly_avg_neu.index.astype(str), monthly_avg_neu.values, marker='^', color='#2ca02c', label='Neutral')
plt.ylabel('Number of Posts', fontsize=15) # Add titles and tags
plt.xticks(rotation=45, fontsize=12) # Custom scale and font size
plt.yticks(fontsize=12)
plt.grid(True, linestyle='--', alpha=0.7) # add grid
plt.legend(fontsize=12) # add legend
plt.tight_layout()
plt.show()
```

d.Topic Sentiment analysis

```
# Filter data by period
pre_pandemic_data = vader_submissions[vader_submissions['Period'] == 'pre_pandemic']
dur_pandemic_data = vader_submissions[vader_submissions['Period'] == 'dur_pandemic']

# Function to calculate sentiment proportions
def calculate_sentiment_proportions(df):
    sentiment_counts = df.groupby('Topic')['Sentiment'].value_counts(normalize=True).unstack(fill_value=0)
    sentiment_proportions = sentiment_counts * 100
    return sentiment_proportions

# Calculate proportions for each period
pre_pandemic_sentiments = calculate_sentiment_proportions(pre_pandemic_data)
dur_pandemic_sentiments = calculate_sentiment_proportions(dur_pandemic_data)
```

```
# Visulize the sentiment polarity by topics
# custom the topic orders
topics = [2, 11, 9, 1, 10, 12, 3, 14, 4, 15, 5, 6, 13, 8, 7, 0]
pre_pandemic_sentiments_ordered = pre_pandemic_sentiments.loc[topics]
dur_pandemic_sentiments_ordered = dur_pandemic_sentiments.loc[topics]

# Extract topics and sentiment proportions from the dataframes
pre_pandemic_positive = pre_pandemic_sentiments_ordered['positive'].tolist()
pre_pandemic_negative = pre_pandemic_sentiments_ordered['negative'].tolist()
pre_pandemic_neutral = pre_pandemic_sentiments_ordered['neutral'].tolist()

dur_pandemic_positive = dur_pandemic_sentiments_ordered['positive'].tolist()
dur_pandemic_negative = dur_pandemic_sentiments_ordered['negative'].tolist()
dur_pandemic_neutral = dur_pandemic_sentiments_ordered['neutral'].tolist()

# Define bar width and positions
bar_width = 0.2  # Width of each bar
gap_width = 0.8  # Gap width between topic groups
r = np.arange(len(topics)) * (3 * bar_width + gap_width)

# Create the plot
fig, ax = plt.subplots(figsize=(16, 8))

# Plot pre-pandemic and during-pandemic positive, negative, and neutral sentiment proportions
ax.bar(r - 2.5 * bar_width, pre_pandemic_positive, color='#ff7f0e', width=bar_width, label='Pre-pandemic Positive')
ax.bar(r - 1.5 * bar_width, dur_pandemic_positive, color='#ffbb78', width=bar_width, label='Dur-pandemic Positive')
ax.bar(r - 0.5 * bar_width, pre_pandemic_negative, color='#1f77b4', width=bar_width, label='Pre-pandemic Negative')
ax.bar(r + 0.5 * bar_width, dur_pandemic_negative, color='#aec7e8', width=bar_width, label='Dur-pandemic Negative')
ax.bar(r + 1.5 * bar_width, pre_pandemic_neutral, color='#2ca02c', width=bar_width, label='Pre-pandemic Neutral')
ax.bar(r + 2.5 * bar_width, dur_pandemic_neutral, color='#98df8a', width=bar_width, label='Dur-pandemic Neutral')

# Add labels and title
x_labels = ['T2', 'T11', 'T9', 'T1', 'T10', 'T12', 'T3', 'T14', 'T4', 'T15', 'T5', 'T6', 'T13', 'T8', 'T7', 'T0'] # rename x label name
ax.set_xticks(r)
ax.set_xticklabels(x_labels, fontsize=13)
ax.set_ylabel('Percentage (%)', fontsize=15)

# Add legend
handles, labels = ax.get_legend_handles_labels()
by_label = dict(zip(labels, handles))
ax.legend(by_label.values(), by_label.keys(), loc='upper left', bbox_to_anchor=(1, 1))

# Display the plot
plt.tight_layout()
plt.show()

# Z-test

# Define the number of posts before and during the pandemic
n_pre_pandemic = 16242  # pre-pandemic posts
n_dur_pandemic = 19139  # dur-pandemic posts

results = [] # create en empty list to store the result

# Loop to calculate the positive and negative sentiment proportions for each topic
for topic in range(pre_pandemic_sentiments.shape[0]):
    # calulate the positive and negative proportion before the pandemic
    pre_positive_prop = pre_pandemic_sentiments.loc[topic, 'positive'] / 100
    pre_negative_prop = pre_pandemic_sentiments.loc[topic, 'negative'] / 100

    # calulate the positive and negative proportion during the pandemic
```

```
    dur_positive_prop = dur_pandemic_sentiments.loc[topic, 'positive'] / 100
    dur_negative_prop = dur_pandemic_sentiments.loc[topic, 'negative'] / 100

    # calulate the number of successes
    count_positive = [pre_positive_prop * n_pre_pandemic, dur_positive_prop * n_dur_pandemic]
    count_negative = [pre_negative_prop * n_pre_pandemic, dur_negative_prop * n_dur_pandemic]

    # the total posts in each period
    nobs = [n_pre_pandemic, n_dur_pandemic]

    # peform the Z-test
    z_stat_positive, p_value_positive = proportions_ztest(count_positive, nobs)
    z_stat_negative, p_value_negative = proportions_ztest(count_negative, nobs)

    # save the result
    results.append({
        'Topic': topic,
        'Positive Z-statistic': z_stat_positive,
        'Positive P-value': p_value_positive,
        'Negative Z-statistic': z_stat_negative,
        'Negative P-value': p_value_negative
    })
```

5.Sentiment Analysis (NRC-EIL)

<u>a. Analyze 8 emotions by NRC-EIL on each post.</u>

```
# Load emotion lexicon and submissions data
lex = pd.read_csv('/Users/arptns/Desktop/Reddit data/NRC-Emotion-Lexicon-Wordlevel-v0.92.txt', sep='\t', names=['word',
'emotion', 'present'])
lex = lex[lex.present == 1]

# NRC-EIL sentiment analysis
def analyze_nrc_sentiment(text):
    sentiment_scores = {emotion: 0 for emotion in lex.emotion.unique()}
    for word in text.split():
        matches = lex[(lex.word == word) & (lex.emotion != 'positive') & (lex.emotion != 'negative')]  # Exclude positive and
negative emotions
        for _, match in matches.iterrows():
            sentiment_scores[match['emotion']] += 1
    return sentiment_scores

nrc_submissions['Emotion'] = nrc_submissions['Selftext'].apply(analyze_nre_sentiment)

# Function to safely evaluate the string representation of the dictionary
def safe_literal_eval(val):
    try:
        return ast.literal_eval(val)
    except (ValueError, SyntaxError):
        return None

# Apply the safe_literal_eval function to the 'Emotion' column
nrc_submissions['Emotion'] = nrc_submissions['Emotion'].apply(safe_literal_eval)

# Transform dictionary into columns
nrc_emotion = nrc_submissions['Emotion'].apply(pd.Series)

# drop the original 'Emotion' column and add emtions to new columns
nrc_submissions = pd.concat([nrc_submissions.drop(['Emotion'], axis=1), nrc_emotion], axis=1)
```

```
# Remove positive and negative colmun
nrc_submissions['Creation Time'] = pd.to_datetime(nrc_submissions['Creation Time'])
nrc_submissions = nrc_submissions.drop(columns=['positive', 'negative'] )
```

b. NRC-EIL descriptive statistics.

```
# Extract a subset of data before and during the pandemic
pre_pandemic_data = nrc_submissions[nrc_submissions['Period'] == 'pre_pandemic']
dur_pandemic_data = nrc_submissions[nrc_submissions['Period'] == 'dur_pandemic']

# Calculate the sum of each emotion before the pandemic
pre_pandemic_emotions_sum = pre_pandemic_data[['trust', 'fear', 'sadness', 'anger', 'surprise', 'disgust', 'joy', 'anticipation']].sum()

# Calculate the proportion of each emotion before the pandemic
pre_pandemic_emotions_percentage = pre_pandemic_emotions_sum / pre_pandemic_emotions_sum.sum() * 100

# Calculate the sum of each emotion during the pandemic
dur_pandemic_emotions_sum = dur_pandemic_data[['trust', 'fear', 'sadness', 'anger', 'surprise', 'disgust', 'joy', 'anticipation']].sum()

# Calculate the proportion of each emotion during the pandemic
dur_pandemic_emotions_percentage = dur_pandemic_emotions_sum / dur_pandemic_emotions_sum.sum() * 100
```

c. Visualize emotion changes over time.

```
# Calculate the sum of the sentiment scores for each row
nrc_submissions['total_emotion'] = nrc_submissions[['trust', 'fear', 'sadness', 'anger', 'surprise', 'disgust', 'joy',
'anticipation']].sum(axis=1)

# Calculate the percentage of each emotion category
emotion_columns = ['trust', 'fear', 'sadness', 'anger', 'surprise', 'disgust', 'joy', 'anticipation']
for emotion in emotion_columns:
    nrc_submissions[f'{emotion}_percent'] = (nrc_submissions[emotion] / nrc_submissions['total_emotion']) * 100

nrc_submissions[['Creation Time', 'trust_percent', 'fear_percent', 'sadness_percent', 'anger_percent', 'surprise_percent',
'disgust_percent', 'joy_percent', 'anticipation_percent']].head()

# Convert 'Creation Time' column to datetime type
nrc_submissions['Creation Time'] = pd.to_datetime(nrc_submissions['Creation Time'])

# Plot the average proportion of joy emotion words
monthly_joy = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['joy_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_joy.index.astype(str), monthly_joy.values, marker='o', color='#ff7f0e', label='Joy', linewidth=3)
plt.xlabel('(a)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()

# Plot the average proportion of sadness emotion words
monthly_sadness = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['sadness_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_sadness.index.astype(str), monthly_sadness.values, marker='o', color='#1f77b4', label='Sadness', linewidth=3)
plt.xlabel('(b)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
```

```
plt.show()

# Plot the average proportion of trust emotion words
monthly_trust = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['trust_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_trust.index.astype(str), monthly_trust.values, marker='o', color='#ff7f0e', label='Trust', linewidth=3)
plt.xlabel('(c)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()

# Plot the average proportion of disgust emotion words
monthly_disgust = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['disgust_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_disgust.index.astype(str), monthly_disgust.values, marker='o', color='#1f77b4', label='Disgust', linewidth=3)
plt.xlabel('(d)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()

# Plot the average proportion of anticipation emotion words
monthly_anticipate = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['anticipation_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_anticipate.index.astype(str), monthly_anticipate.values, marker='o', color='#ff7f0e', label='Anticipation',
linewidth=3)
plt.xlabel('(e)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()

# Plot the average proportion of fear emotion words
monthly_fear = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['fear_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_fear.index.astype(str), monthly_fear.values, marker='o', color='#1f77b4', label='Fear', linewidth=3)
plt.xlabel('(f)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()

# Plot the average proportion of surprise emotion words
monthly_surprise = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['surprise_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_surprise.index.astype(str), monthly_surprise.values, marker='o', color='#ff7f0e', label='Surprise', linewidth=3)
plt.xlabel('(g)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()
```

```
# Plot the average proportion of anger emotion words
monthly_anger = nrc_submissions.groupby(nrc_submissions['Creation Time'].dt.to_period('M'))['anger_percent'].mean()
plt.figure(figsize=(20, 10))
plt.plot(monthly_anger.index.astype(str), monthly_anger.values, marker='o', color='#1f77b4', label='Anger', linewidth=3)
plt.xlabel('(h)', fontsize=30)
plt.ylabel('Percentage (%)', fontsize=30)
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.legend(fontsize=30)
plt.show()
```

d.Topic Emotion analysis

```
# Step 1: Aggregate data by Period and Topic, summing the emotion categories
agg_nrc = nrc_submissions.groupby(['Period', 'Topic']).agg({
    'joy': 'sum',
    'trust': 'sum',
    'anticipation': 'sum',
    'surprise': 'sum',
    'fear': 'sum',
    'sadness': 'sum',
    'disgust': 'sum',
    'anger': 'sum'
}).reset_index()

# Step 2: Calculate the total number of emotions per topic to use for proportion calculations
agg_nrc['total_emotions'] = agg_nrc[['joy', 'trust', 'anticipation', 'surprise', 'fear', 'sadness', 'disgust', 'anger']].sum(axis=1)

# Step 3: Calculate proportions for each emotion so that each topic sums to 100%
for emotion in ['joy', 'trust', 'anticipation', 'surprise', 'fear', 'sadness', 'disgust', 'anger']:
    agg_nrc[emotion + '_prop'] = (agg_nrc[emotion] / agg_nrc['total_emotions']) * 100

# Step 4: Split the data into two separate tables
pre_pandemic_nrc = agg_nrc[agg_nrc['Period'] == 'pre_pandemic']
dur_pandemic_nrc = agg_nrc[agg_nrc['Period'] == 'dur_pandemic']

# Step 5: Select the final columns for display, including both raw counts and proportions
emotion_columns = ['joy', 'trust', 'anticipation', 'surprise', 'fear', 'sadness', 'disgust', 'anger']
pre_pandemic_nrc = pre_pandemic_nrc[['Topic'] + emotion_columns +
                                    [emotion + '_prop' for emotion in emotion_columns]]

dur_pandemic_nrc = dur_pandemic_nrc[['Topic'] + emotion_columns +
                                    [emotion + '_prop' for emotion in emotion_columns]]

# Visualize the emotion distribution for each period
# custom the topic order according the themes
topics = [2, 11, 9, 1, 10, 12, 3, 14, 4, 15, 5, 6, 13, 8, 7, 0]
x_labels = ['T2', 'T11', 'T9', 'T1', 'T10', 'T12', 'T3', 'T14', 'T4', 'T15', 'T5', 'T6', 'T13', 'T8', 'T7', 'T0']
pre_pandemic_sorted = pre_pandemic_nrc.set_index('Topic').loc[topics].reset_index()
dur_pandemic_sorted = dur_pandemic_nrc.set_index('Topic').loc[topics].reset_index()

# extract eight emotions before the pandemic
pre_pandemic = {
    'joy': pre_pandemic_sorted['joy_prop'].tolist(),
    'trust': pre_pandemic_sorted['trust_prop'].tolist(),
    'anticipation': pre_pandemic_sorted['anticipation_prop'].tolist(),
    'surprise': pre_pandemic_sorted['surprise_prop'].tolist(),
    'fear': pre_pandemic_sorted['fear_prop'].tolist(),
    'sadness': pre_pandemic_sorted['sadness_prop'].tolist(),
    'disgust': pre_pandemic_sorted['disgust_prop'].tolist(),
```

```
    'anger': pre_pandemic_sorted['anger_prop'].tolist()
}

# extract eight emotions during the pandemic
dur_pandemic = {
    'joy': dur_pandemic_sorted['joy_prop'].tolist(),
    'trust': dur_pandemic_sorted['trust_prop'].tolist(),
    'anticipation': dur_pandemic_sorted['anticipation_prop'].tolist(),
    'surprise': dur_pandemic_sorted['surprise_prop'].tolist(),
    'fear': dur_pandemic_sorted['fear_prop'].tolist(),
    'sadness': dur_pandemic_sorted['sadness_prop'].tolist(),
    'disgust': dur_pandemic_sorted['disgust_prop'].tolist(),
    'anger': dur_pandemic_sorted['anger_prop'].tolist()
}

def plot_emotion_distribution(data, topics, x_labels):
    bar_width = 0.5  # set the width of the bar
    r = np.arange(len(topics))

    # set the color to orange (postive emotions) and blue (negative emotions)
    positive_colors = ['#ffcc99', '#ff9933', '#ff7f0e', '#ff6f00']
    negative_colors = ['#aec7e8', '#6699cc', '#1f77b4', '#0f5292']

    fig, ax = plt.subplots(figsize=(16, 8))
    bottoms = np.zeros(len(r))

    # plot the emotion stacked chart
    for i, (emotion, values) in enumerate(data.items()):
        bars = ax.bar(r, values, bottom=bottoms, width=bar_width, label=emotion.capitalize(), color=positive_colors[i] if i < 4 else
negative_colors[i-4])
        # add the proportion text into bars
        for bar, value in zip(bars, values):
            ax.text(bar.get_x() + bar.get_width() / 2, bar.get_y() + bar.get_height() / 2, f'{value:.1f}%', ha='center', va='center',
fontsize=8)
        bottoms += np.array(values)

    # add labels and title
    ax.set_xticks(r)
    ax.set_xticklabels(x_labels, fontsize=13)
    ax.set_ylabel('Proportion (%)', fontsize=15)
    ax.legend(loc='upper left', bbox_to_anchor=(1, 1))

    plt.tight_layout()
    plt.show()

# plot the emotions distribution before the pandemic
plot_emotion_distribution(pre_pandemic, topics, x_labels)

# plot the emotions distribution during the pandemic
plot_emotion_distribution(dur_pandemic, topics, x_labels)
```

Reference

Almerekhi, H., Jansen, S. B. B. J., & Kwak, C. B. H. (2020). *Investigating toxicity across multiple Reddit communities, users, and moderators.* Companion Proceedings of the Web Conference 2020. https://doi.org/10.1145/3366424.3382091

American Psychological Association. (2021). One year later, a new wave of pandemic health concerns. Retrieved July 08 2024 from https://www.apa.org/news/press/releases/stress/2021/one-year-pandemic-stress

Almandoz, J. P., Xie, L., Schellinger, J. N., Mathew, M. S., Marroquin, E. M., Murvelashvili, N., Khatiwada, S., Kukreja, S., McAdams, C., & Messiah, S. E. (2022). *Changes in body weight, health behaviors, and mental health in adults with obesity during the COVID-19 pandemic.* Obesity (Silver Spring, Md.), 30(9), 1875–1886. https://doi.org/10.1002/oby.23501

AI Zaman, K. A., Ahmed, S., Alshamsi, A., Alshamsi, A., Alshdaifat, B., Alaleeli, S., & Mussa, B. M. (2023). *Impact of COVID-19 pandemic on weight change among adults in the UAE.* International Journal of General Medicine, Volume 16, 1661–1670. https://doi.org/10.2147/ijgm.s407934

Alipour, S., Galeazzi, A., Sangiorgio, E., Avalle, M., Bojic, L., Cinelli, M., & Quattrociocchi, W. (2024). *Cross-platform social dynamics: an analysis of ChatGPT and COVID-19 vaccine conversations.* Scientific Reports, 14(1), 2789. https://doi.org/10.1038/s41598-024-53124-x

Barsade, S. G. (2002). *The ripple effect: Emotional contagion and its influence on group behavior.* Administrative science quarterly, 47(4), 644-675. https://doi.org/10.2307/3094912

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet allocation.* Journal of Machine Learning Research, 3(Jan), 993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Blei, D. M., & Lafferty, J. D. (2006, June). *Dynamic topic models.* In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). https://doi.org/10.1145/1143844.1143859

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). *Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model.* Multivariate behavioral research, 53(3), 293-314. https://doi.org/10.1080/00273171.2018.1439722

Brand, R., Timme, S., & Nosrat, S. (2020). *When pandemic hits: exercise frequency and subjective well-being during covid-19 pandemic.* Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.570567

Bhutani, S., & Cooper, J. A. (2020). *COVID‐19–Related Home Confinement in Adults: Weight Gain Risks and Opportunities.* Obesity (Silver Spring, Md.), 28(9), 1576. https://doi.org/10.1002/oby.22904

Bennett, G., Young, E., Butler, I., & Coe, S. (2021). *The impact of lockdown during the COVID-19 outbreak on dietary habits in various population groups: a scoping review.* Frontiers in nutrition, 8, 626432. https://doi.org/10.3389/fnut.2021.626432

Burdett, A., Davillas, A., & Etheridge, B. (2021). *Weather, mental health, and mobility during the first wave of the covid‐19 pandemic.* Health Economics, 30(9), 2296-2306. https://doi.org/10.1002/hec.4371

Barenbaum, S. R., Saunders, K. H., Chan, K. M., Crowley, W. J., Redmond, I. P., Casper, A. J., ... & Shukla, A. P. (2022). *Medical weight management protects against weight gain during the COVID‐19 pandemic.* Obesity Science & Practice, 8(5), 682-687. https://doi.org/10.1002/osp4.601

Burnatowska, E., Surma, S., & Olszanecka-Glinianowicz, M. (2022). *Relationship between mental health and emotional eating during the COVID-19 pandemic: A systematic review.* Nutrients, 14(19), 3989. https://doi.org/10.3390/nu14193989

Britt, R. K., Franco, C. L., & Jones, N. (2023). *Trends and challenges within Reddit and health communication research: A systematic review.* Communication and the Public, 8(4), 402-417. https://doi.org/10.1177/20570473231209075

Cunha, T., Weber, I., & Pappa, G. (2017, April). *A warm welcome matters! the link between social feedback and weight loss in/r/loseit*. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1063-1072). https://doi.org/10.1145/3041021.3055131

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., ... & Scala, A. (2020). *The COVID-19 social media infodemic.* Scientific Reports, 10(1), 16598. https://doi.org/10.1038/s41598-020-73510-5

Chan, A. K., Nickson, C. P., Rudolph, J. W., Lee, A., & Joynt, G. M. (2020). *Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic.* Anaesthesia, 75(12), 1579. https://doi.org/10.1111/anae.15057

Chen, P., Mao, L., Nassis, G. P., Harmer, P., Ainsworth, B. E., & Li, F. (2020). *Coronavirus disease (COVID-19): The need to maintain regular physical activity while taking precautions.* Journal of sport and health science, 9(2), 103. https://doi.org/10.1016/j.jshs.2020.02.001

Cecchetto, C., Aiello, M., Gentili, C., Ionta, S., & Osimo, S. A. (2021). *Increased emotional eating during COVID-19 associated with lockdown, psychological and social distress.* Appetite, 160, 105122. https://doi.org/10.1016/j.appet.2021.105122

Cheval, B., Sivaramakrishnan, H., Maltagliati, S., Fessler, L., Forestier, C., Sarrazin, P., Orsholits, D., Chalabaev, A., Sander, D., Ntoumanis, N., & Boisgontier, M. P. (2021). *Relationships between changes in self-reported physical activity, sedentary behaviour and health during the coronavirus (COVID-19) pandemic in France and Switzerland.* Journal of sports sciences, 39(6), 699–704. https://doi.org/10.1080/02640414.2020.1841396

Caldwell, A. E., Thomas, E. A., Rynders, C., Holliman, B. D., Perreira, C., Ostendorf, D. M., & Catenacci, V. A. (2022). *Improving lifestyle obesity treatment during the COVID-19 pandemic and beyond: new challenges for weight management.* Obesity science & practice, 8(1), 32-44. https://doi.org/10.1002/osp4.540

Chi, Y., & Chen, H. Y. (2023). *Investigating substance use via Reddit: systematic scoping review*. Journal of Medical Internet Research, 25, e48905. https://doi.org/10.2196/48905

Chen, K., He, Z., Burghardt, K., Zhang, J., & Lerman, K. (2024). *IsamasRed: A Public Dataset Tracking Reddit Discussions on Israel-Hamas Conflict.* arXiv preprint arXiv:2401.08202. https://doi.org/10.48550/arXiv.2401.08202

Dhankar, A., & Katz, A. (2023). *Tracking pregnant women's mental health through social media: an analysis of reddit posts.* JAMIA open, 6(4), ooad094. https://doi.org/10.1093/jamiaopen/ooad094

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals.* John Murray. https://doi.org/10.1037/10001-000

Renzo, L., Gualtieri, P., Pivari, F., Soldati, L., Attinà, A., Cinelli, G., Leggeri, C., Caparello, G., Barrea, L., Scerbo, F., Esposito, E., & De Lorenzo, A. (2020). *Eating habits and lifestyle changes during COVID-19 lockdown: an Italian survey.* Journal of translational medicine, 18(1), 229. https://doi.org/10.1186/s12967-020-02399-5

Ferrara, E., & Yang, Z. (2015). *Measuring emotional contagion in social media.* PloS one, 10(11), e0142390. https://doi.org/10.1371/journal.pone.0142390

Fan R., Xu K., Zhao J. (2018*). An agent-based model for emotion contagion and competition in online social media.* Physica a: Statistical mechanics and its applications, 495, 245–259. https://doi.org/10.1016/j.physa.2017.12.086

Flint, S. W. (2020). *Stigmatizing media portrayal of obesity during the coronavirus (COVID-19) pandemic.* Frontiers in Psychology, 11, 2124. https://doi.org/10.3389/fpsyg.2020.02124

Grimmer, J., & Stewart, B.M. 2017. *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.* Polit Anal, 21(3): 267-297. https://doi.org/10.1093/pan/mps028

Gaudette, T., Scrivens, R., Davies, G., & Frank, R. (2021). *Upvoting extremism: Collective identity formation and the extreme right on Reddit.* New Media & Society, 23(12), 3491-3508. https://doi.org/10.1177/1461444820958123

González-Monroy, C., Gómez-Gómez, I., Olarte-Sánchez, C. M., & Motrico, E. (2021). *Eating behaviour changes during the COVID-19 pandemic: a systematic review of longitudinal studies.* International journal of environmental research and public health, 18(21), 11130. https://doi.org/10.3390/ijerph182111130

Glazer, S. A., & Vallis, M. (2022). *Weight gain, weight management and medical care for individuals living with overweight and obesity during the COVID-19 pandemic (EPOCH Study).* Obesity Science & Practice, 8(5), 556-568. https://doi.org/10.1002/osp4.591

Google Trends. (2024). *Weight loss - Explore - Google Trends.* Retrieved July 08, 2024, from https://trends.google.com/trends/explore?date=all&q=weight%20loss

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). *Emotional contagion.* Current directions in psychological science, 2(3), 96-100. https://doi.org /10.1111/1467-8721.ep10770953.

Hutto, C. J., & Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text.* Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf.

Hutto, C. and Gilbert, É. (2014). *vader: a parsimonious rule-based model for sentiment analysis of social media text.* Proceedings of the International Aaai Conference on Web and social media, 8(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Highfield, T., & Leaver, T. (2016). *Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji.* Communication Research and Practice, 2(1), 47-62. https://doi.org/10.1080/22041451.2016.1155332

Hale, T., Angrist, N., Kira, B., Petherick, A., Phillips, T., & Webster, S. (2020). Variation in government responses to COVID-19. https://ora.ox.ac.uk/objects/uuid:0ab73a02-ca18-4e1f-a41b-cfeea2d30e81/files/s3j3332380

Hwang, Y., Kim, H. J., Choi, H. G., & Lee, J. (2020). *Exploring abnormal behavior patterns of online users with emotional eating behavior: topic modeling study.* Journal of Medical Internet Research, 22(3), e15700. https://doi.org/10.2196/15700

Jane, M., Hagger, M., Foster, J., Ho, S., & Pal, S. (2018). *Social media for health promotion and weight management: a critical debate.* BMC public health, 18, 1-7. https://doi.org/10.2196/15700

Janßen, M., Chang, B., Hristov, H., Pravst, I., Profeta, A., & Millard, J. (2021). *Changes in food consumption during the covid-19 pandemic: analysis of consumer survey data from the first lockdown period in denmark, germany, and slovenia.* Frontiers in Nutrition, 8. https://doi.org/10.3389/fnut.2021.635859

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). *Experimental evidence of massive-scale emotional contagion through social networks.* Proceedings of the National Academy of Sciences, 111(24), 8788-8790. https://doi.org/10.1073/pnas.1320040111

Kluyver, Thomas, Ragan-Kelley, Benjamin, Pérez, Fernando, Granger, Brian, Bussonnier, Matthias, Frederic, Jonathan, Kelley, Kyle, Hamrick, Jessica, Grout, Jason, Corlay, Sylvain, Ivanov, Paul, Avila, Damián, Abdalla, Safia, Willing, Carol and Jupyter development team, (2016) *Jupyter Notebooks – a publishing format for reproducible computational workflows.* Loizides, Fernando and Scmidt, Birgit (eds.) In Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. pp. 87-90 . http://doi.org/10.3233/978-1-61499-649-1-87

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). *A multivariate statistical model for emotion dynamics.* Emotion, 18(5), 739. http://doi.org/10.1037/emo0000384

Kozicki, B., & Gornikiewicz, M. (2020). *Unemployment rate in Poland and USA during COVID-19 pandemic: a case study.* http://doi.org/10.35808/ersj/1861

Keel, P. K., Gomez, M. M., Harris, L., Kennedy, G. A., Ribeiro, J., & Joiner, T. E. (2020). *Gaining "The Quarantine 15:" Perceived versus observed weight changes in college students in the wake of COVID-19.* International Journal of Eating Disorders, 53(11), 1801-1808. http://doi.org/10.1002/eat.23375

Knell, G., Robertson, M. C., Dooley, E. E., Burford, K., & Mendez, K. S. (2020). *Health behavior changes during COVID-19 pandemic and subsequent "stay-at-home" orders*. International journal of environmental research and public health, 17(17), 6268. http://doi.org/10.3390/ijerph17176268

Kuk, J. L., Christensen, R. A., Kamran Samani, E., & Wharton, S. (2021). *Predictors of Weight Loss and Weight Gain in Weight Management Patients during the COVID-19 Pandemic.* Journal of Obesity, 2021(1), 4881430. http://doi.org/10.1155/2021/4881430

Khubchandani, J., Price, J. H., Sharma, S., Wiblishauser, M. J., & Webb, F. J. (2022). *COVID-19 pandemic and weight gain in American adults: a nationwide population-based study.* Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 16(1), 102392. http://doi.org/10.1016/j.dsx.2022.102392

Lopez-Leon, S., Wegman-Ostrosky, T., Perelman, C., Sepulveda, R., Rebolledo, P. A., Cuapio, A., & Villapol, S. (2021). *More than 50 long-term effects of COVID-19: a systematic review and meta-analysis.* Scientific reports, 11(1), 1-12. https://doi.org/10.1101/2021.01.27.21250617

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). *Language and the perception of emotion.* Emotion, 6(1), 125. https://doi.org/10.1037/1528-3542.6.1.125

Lazarus, R. S. (1991). *Emotion and adaptation (Vol. 557).* Oxford University Press.

Lu, Y., Hu, X., Wang, F., Kumar, S., Liu, H., & Maciejewski, R. (2015, May). *Visualizing social media sentiment in disaster scenarios.* In Proceedings of the 24th international conference on world wide web (pp. 1211-1215). https://doi.org/10.1145/2740908.2741720

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

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). *Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study*. Journal of medical Internet research, 22(10), e22635. https://doi.org/10.2196/22635

Liu, Y. and Yin, Z. (2020). *Understanding weight loss via online discussions: content analysis of reddit posts using topic modeling and word clustering techniques.* Journal of Medical Internet Research, 22(6), e13745. https://doi.org/10.2196/13745

Massanari, A. L. (2015). *#Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures.* New Media & Society, 19(3), 329-346. https://doi.org/10.1177/1461444815608807

Mohan, S., Guha, A., Harris, M., Popowich, F., Schuster, A., & Priebe, C. (2017). *The impact of toxic language on the health of reddit communities.* In Advances in Artificial Intelligence: 30th Canadian Conference on Artificial Intelligence, Canadian AI 2017, Edmonton, AB, Canada, May 16-19, 2017, Proceedings 30 (pp. 51-56). Springer International Publishing. https://doi.org/10.1007/978-3-319-57351-9_6

Mohammad, S. (2018). *Word Affect Intensities.* In Proceedings of the 11th edition of the Language Resources and Evaluation Conference, LREC 2018. https://doi.org/10.48550/arXiv.1704.08798

Miao, Wanwan., Guiping, Zhang., Yu, Bai., Dongfeng, Cai. (2019). *Improving Accuracy of Key Information Acquisition for Social Media Text Summarization.* https://doi.org/10.1109/IUCC/DSCI/SMARTCNS.2019.00094

McKibbin, W. and Fernando, R. (2020). *Global macroeconomic scenarios of the covid-19 pandemic.* SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3635103

Muscogiuri, G., Barrea, L., Savastano, S., & Colao, A. (2020). *Nutritional recommendations for CoVID-19 quarantine.* European journal of clinical nutrition, 74(6), 850-851. https://doi.org/10.1038/s41430-020-0635-2

Mason, T. B., Barrington-Trimis, J., & Leventhal, A. M. (2021*). Eating to cope with the COVID-19 pandemic and body weight change in young adults.* Journal of Adolescent Health, 68(2), 277-283. https://doi.org/10.1016/j.jadohealth.2020.11.011

Manova, V., Grosso, F., Khoury, B., & Pagnini, F. (2024). Social anxiety: topics and emotions shared on Reddit before and during the coronavirus pandemic. Current Psychology, 1-10. https://doi.org/10.1007/s12144-024-05891-z

Nabity-Grover, T., Cheung, C. M. K., & Thatcher, J. B. (2020). *Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media.* International journal of information management, 55, 102188. https://doi.org/10.1016/j.ijinfomgt.2020.102188

Naskar, D., Singh, S. R., Kumar, D., Nandi, S., & Rivaherrera, E. O. D. L. (2020). *Emotion dynamics of public opinions on twitter.* ACM Transactions on Information Systems (TOIS), 38(2), 1-24. https://doi.org/10.1145/3379340

Onyeaka, H., Anumudu, C. K., Al-Sharify, Z. T., Egele-Godswill, E., & Mbaegbu, P. (2021). *COVID-19 pandemic: A review of the global lockdown and its far-reaching effects.* Science progress, 104(2), 00368504211019854. https://doi.org/10.1177/00368504211019854

Plutchik, R. (1980). *A general psychoevolutionary theory of emotion.* Emotion: Theory, research, and experience, 1. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis.* Foundations and Trends® in information retrieval, 2(1–2), 1-135. http://doi.org/10.1561/1500000001

Palmeira, A. L., Branco, T. L., Martins, S. C., Minderico, C. S., Silva, M. N., Vieira, P. N., ... & Teixeira, P. J. (2010). *Change in body image and psychological well-being during behavioral obesity treatment: Associations with weight loss and maintenance.* Body Image, 7(3), 187-193. https://doi.org/10.1016/j.bodyim.2010.03.002

Papacharissi, Z. (2014). *Affective publics: Sentiment, technology, and politics.* Oxford Studies in Digital Politics. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001

Patel, R., Chang, T., Greysen, S. R., & Chopra, V. (2015). *Social media use in chronic disease: a systematic review and novel taxonomy.* The American journal of medicine, 128(12), 1335-1350. https://doi.org/10.1016/j.amjmed.2015.06.015

Papacharissi, Z. (2016). *Affective publics and structures of storytelling: Sentiment, events and mediality.* Information, communication & society, 19(3), 307-324. https://doi,org/10.1080/1369118X.2015.1109697

Pappa, G., Cunha, T., Bicalho, P., Ribeiro, A., Silva, A., Meira, W., … & Du, H. (2017). *Factors associated with weight change in online weight management communities: a case study in the loseit reddit community.* Journal of Medical Internet Research, 19(1), e17. https://doi.org/10.2196/jmir.5816

Primack, B. A., Shensa, A., Escobar-Viera, C. G., Barrett, E. L., Sidani, J. E., Colditz, J. B., & James, A. E. (2017). *Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among US young adults.* Computers in human behavior, 69, 1-9. https://doi.org/10.1016/j.chb.2016.11.013

Pearl, R. L. (2020). *Weight stigma and the "Quarantine-15".* Obesity (Silver Spring, Md.), 28(7), 1180. https://doi.org/10.1002/oby.22850

Pellegrini, C. A., Webster, J., Hahn, K. R., Leblond, T. L., & Unick, J. L. (2021). *Relationship between stress and weight management behaviors during the COVID-19 pandemic among those enrolled in an internet program.* Obesity science & practice, 7(1), 129-134. https://doi.org/10.1002/osp4.465

Röder, M., Both, A., & Hinneburg, A. (2015, February). *Exploring the space of topic coherence measures.* In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408). https://doi.org/10.1145/2684822.2685324

Renzo, L., Gualtieri, P., Pivari, F., Soldati, L., Attinà, A., Cinelli, G., Leggeri, C., Caparello, G., Barrea, L., Scerbo, F., Esposito, E., & De Lorenzo, A. (2020). *Eating habits and lifestyle changes during COVID-19 lockdown: an Italian survey.* Journal of translational medicine, 18(1), 229. https://doi.org/10.1186/s12967-020-02399-5

Robinson, E., Boyland, E., Chisholm, A., Harrold, J., Maloney, N. G., Marty, L., ... & Hardman, C. A. (2021). *Obesity, eating behavior and physical activity during COVID-19 lockdown: A study of UK adults.* Appetite, 156, 104853. https://doi.org/10.1016/j.appet.2020.104853

Rosenkilde, S., Sørensen, T. I., Algren, M. H., & Thygesen, L. C. (2024). *Changes in weight status during the COVID-19 pandemic: impact of educational level and mental health.* European Journal of Public Health, 34(1), 190-195. https://doi.org/10.1093/eurpub/ckad188

Reddit. (n.d.). *r/loseit: Lose weight, get in shape!.* Retrieved Mar 01 2024, from https://www.reddit.com/r/loseit/

Reddit. (n.d.). *API Documentation.* Retrieved Mar 15 2024, from https://www.reddit.com/dev/api.

Reddit. (n.d.). *API Terms.* Retrieved Mar 15 2024, from https://www.reddit.com/wiki/api-terms/.

Reddit. (n.d.). *Reddit Data API Wiki.* Retrieved Mar 15 2024, from https://support.reddithelp.com/hc/en-us/articles/16160319875092.

Swinburn, B. A., Caterson, I., Seidell, J. C., & James, W. P. (2004). *Diet, nutrition and the prevention of excess weight gain and obesity.* Public health nutrition, 7(1A), 123–146. https://doi.org/10.1079/phn2003585

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). *Can emotions be truly group level? Evidence regarding four conceptual criteria.* Journal of personality and social psychology, 93(3), 431. https://doi.org/10.1037/0022-3514.93.3.431

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012, July). *Exploring topic coherence over many models and many topics.* In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (pp. 952-961). https://aclanthology.org/D12-1087

Shah, S. G. S., & Farrow, A. (2020). *A commentary on "World Health Organization declares global emergency: A review of the 2019 novel Coronavirus (COVID-19)".* International journal of surgery, 76, 128-129. https://doi.org/10.1016/j.ijsu.2020.03.001

Seifert, A., & Hassler, B. (2020). *Impact of the COVID-19 pandemic on loneliness among older adults.* Frontiers in sociology, 5, 590935. https://doi.org/10.3389/fsoc.2020.590935

Saud, M., Mashud, M. I., & Ida, R. (2020). *Usage of social media during the pandemic: Seeking support and awareness about COVID-19 through social media platforms.* Journal of public affairs, 20(4), e2417. https://doi.org/10.1002/pa.2417

Schneider, K. T., & Carpenter, N. J. (2020). *Sharing# MeToo on Twitter: Incidents, coping responses, and social reactions.* Equality, Diversity and Inclusion: An International Journal, 39(1), 87-100. https://doi.org/10.1108/EDI-09-2018-0161

Smith, K. R., Jansen, E., Thapaliya, G., Aghababian, A. H., Chen, L., Sadler, J. R., & Carnell, S. (2021). *The influence of COVID-19-related stress on food motivation.* Appetite, 163, 105233. https://doi.org/10.1016/j.appet.2021.105233

Steinert, S. (2021). *Corona and value change. The role of social media and emotional contagion.* Ethics and Information Technology, 23(Suppl 1), 59-68. https://doi.org/10.1007/s10676-020-09545-z

Schienkiewitz, A., Damerow, S., Richter, A., & Mensink, G. B. (2022). *How has body weight changed since the beginning of the COVID-19 pandemic?*. Journal of Health Monitoring, 7(4), 54. https://doi.org/10.25646/10670

Tison, G. H., Avram, R., Kuhar, P., Abreau, S., Marcus, G. M., Pletcher, M. J., & Olgin, J. E. (2020). *Worldwide effect of COVID-19 on physical activity: a descriptive study.* Annals of internal medicine, 173(9), 767-770. https://doi.org/10.7326/M20-2665

Thelwall, M., & Thelwall, S. (2020). *A thematic analysis of highly retweeted early COVID-19 tweets: consensus, information, dissent and lockdown life.* Aslib Journal of Information Management, 72(6), 945-962. https://doi.org/10.1108/AJIM-05-2020-0134

Tang, H., Kim, S., Laforet, P. E., Tettey, N. S., & Basch, C. H. (2022). *Loss of weight gained during the COVID-19 pandemic: content analysis of YouTube videos.* JMIR Formative Research, 6(2), e35164. https://doi.org/10.2196/35164

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace.

Van der Nagel, E. (2013). *Faceless bodies: Negotiating technological and cultural codes on Reddit Gonewild.* Scan: Journal of Media Arts Culture, 10(2), 1-10.

Vanoh, D., & Juanis, A. W. (2023). *Impact of Covid-19 lockdown on sleep quality, physical activity, and body weight among university students: A retrospective cross-sectional study.* Journal of education and health promotion, 12, 79. https://doi.org/10.4103/jehp.jehp_1381_21

World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19.* Retrieved July 22 2024 from https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

World Health Organization (WHO). *Impact of COVID-19 on people's livelihoods, their health and our food systems.* Retrieved July 22 2024 from https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems

World Health Organization. (2020, March 11). *WHO Director-General's opening remarks at the media briefing on COVID-19.* Retrieved July 22 2024 from https://www.who.int/director-general/speeches/detail/who-director-general-s-openingremarks-at-the-media-briefing-on-covid-19---11-march-2020.

Wanchoo, K., Abrams, M., Merchant, R. M., Ungar, L., & Guntuku, S. C. (2023). *Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19.* Plos one, 18(2), e0280337. https://doi.org/10.1371/journal.pone.0280337

Xia, Y., Zhu, H., Lu, T., Zhang, P., & Gu, N. (2020). *Exploring antecedents and consequences of toxicity in online discussions: A case study on reddit.* Proceedings of the ACM on Human-computer Interaction, 4(CSCW2), 1-23. https://doi.org/10.1145/3415179

Yu, S., Eisenman, D., & Han, Z. (2021). *Temporal dynamics of public emotions during the COVID-19 pandemic at the epicenter of the outbreak: sentiment analysis of Weibo posts from Wuhan.* Journal of medical Internet research, 23(3), e27078. https://doi.org/10.2196/27078

Yang, Q., Ledbetter, A. M., Zhuang, J., & Richards, A. S. (2023). *Theme and sentiment of posts in a weight loss subreddit predict popularity, engagement, and users' weight loss: a computational approach.* Human Communication Research, 49(4), 452-459. https://doi.org/10.1093/hcr/hqad023

Zachary, Z., Brianna, F., Brianna, L., Garrett, P., Jade, W., Alyssa, D., & Mikayla, K. (2020). *Self-quarantine and weight gain related risk factors during the COVID-19 pandemic.* Obesity research & clinical practice, 14(3), 210-216. https://doi.org/10.1016/j.orcp.2020.05.004