ANALYSIS ON EMOTIONS AND PERCEPTIONS RELATED TO COVID19 IN MILAN AND NEW YORK CITY

Maria Duran
mfd6863@uncw.edu
University of North Carolina at Wilmington

Ixchel Peralta
iyp4754@uncw.edu
University of North Carolina at Wilmington

Pawarisa Sears
pis4402@uncw.edu
University of North Carolina at Wilmington

1. INTRODUCTION

In the wake of the COVID19 pandemic, the world has witnessed not only a health crisis but also a significant emotional upheaval. This technical report presents a comprehensive data analysis focused on the emotional impact of COVID19, specifically in the urban landscapes of Milan, Italy, and New York City, USA.

Utilizing a robust dataset from Harvard Dataverse, this analysis delves into the emotional responses and sentiments expressed by individuals during the peak lockdown periods of March and April 2020, and extends to subsequent years. Our data-driven approach aims to decode the emotional fabric of these cities, taking into account variables such as age, gender, ethnicity, and occupation, to paint a detailed picture of the pandemic's psychological footprint.

The primary objectives of this report are to dissect the variety of emotional responses triggered by the pandemic and the associated lockdowns, understand public sentiment towards measures like social distancing and vaccination, and to evaluate the impact on specific demographic groups including students and working professionals. The analysis also includes a comparative study of social comfort levels across different phases of the pandemic. By applying data analysis techniques, this report endeavors to offer a nuanced understanding of the pandemic's emotional impact on diverse population segments.

2. BACKGROUND

2.1 COVID19 Impact on Milan and New York

The COVID19 pandemic significantly impacted Milan, Italy, and New York City, USA, both being among the early and hardest hit areas. Milan, as an epicenter in Italy, saw a surge in cases and a high fatality rate, with its healthcare system overwhelmed [7]. In New York City, the virus's rapid spread led to severe health, social, and economic challenges [14]. The implementation of lockdowns, social distancing, and mask mandates by local authorities were critical measures in managing the pandemic in these urban settings.

2.2 Purpose of Analysis

This data analysis aims to explore the mental health implications of the pandemic. With studies[11][16] indicating increased anxiety, depression, and unmet mental health needs, particularly among those with pre-existing conditions, our analysis focuses on several key aspects:

- Emotional states of people living with others during lockdown.
- Experiences of those living alone, including their interactions and attitudes towards COVID19 regulations.
- Public sentiment towards unvaccinated individuals.
- Emotional impact of testing positive for COVID19.
- Students' feelings about studying during the pandemic.
- Professionals' emotional experiences and work situation changes.
- Comparing comfort levels with social activities between Winter 2021 and Summer 2022.
- Public emotions about returning to pre-pandemic normalcy.
- Main emotions associated with COVID19.
- Analyzing COVID19 related tweets using a GPT model for deeper insight.

2.3 Datasets

The main dataset[2] comprises responses from students and professionals in Milan and New York, collected in mid-2022. Covering a range of topics, from interpersonal relationships to professional and academic situations, it offers a comprehensive view of pandemic life. The survey employed various question formats, including multiple-choice, ranking, and scale-based queries. Our methodology includes data cleaning, categorization, visualization, and comparative analysis to synthesize these insights.

For the secondary dataset[6], we focus on the text content of tweets from various cities,using a GPT model specifically designed to analyze and understand public sentiment about COVID19 as expressed on Twitter.

3. DATA PROCESSING

3.1 Main Data Source, Harvard Dataverse

- Initial Data Structure: The dataset started with 3005 entries spread over 123 columns.
- Column Reduction: To focus on relevant data, 27 columns were dropped, leaving 96 columns for analysis.
- Renaming Columns: For better understanding, 93 columns were renamed.
- Handling Missing Values: Missing values, ranging from 12 to 99 % across columns, were meticulously addressed, with some replaced by "No" to indicate explicit non-responses.
- Data Reorganization: Reorganized columns and tailored the data to suit the study's specific objectives.
- Representation of Results: The dataset's categorical nature presented initial challenges in visual representation.

3.2 Secondary Dataset, "COVID19 Tweets"

3.2.1 Data Overview and Cleaning

- Dataset Overview: Began with 179,108 tweets, with a primary focus on the 'text' column.
- Data Cleaning: Regular expressions were used for cleaning, resulting in the exclusion of 37 % of tweets that lacked direct mentions of "COVID19," "coronavirus," or "covid."

3.2.2 Creating a GPT Assistant

- Tool: A custom GPT[9] sentiment analysis assistant was developed.
- Analysis Process: This assistant was utilized for an in-depth sentiment analysis of the tweet data.
- Iterations of Assistant Improvement: More than 20 updates were applied to enhance the ChatGPT assistant's ability to analyze social media sentiments.
- Video Attachment: An accompanying video is included in the project folder, showcasing the final customization of the ChatGPT assistant. This video provides a detailed walk through of the final version of the assistant, demonstrating its application in analyzing and interpreting the sentiment data from the tweets.
- Example of Initial Prompting:
  - Role: You are an expert data scientist specializing in sentiment analysis.
  - Objective: Conduct a comprehensive sentiment analysis on a dataset containing social media opinions related to COVID19.
  - Instructions: Analyze the 'text' column in the provided dataset, focusing on identifying sentiments expressed about the pandemic. Categorize these sentiments into positive, negative, and neutral. Provide summary statistics and insights into the prevailing public mood and opinions regarding COVID19 as reflected in the tweets.

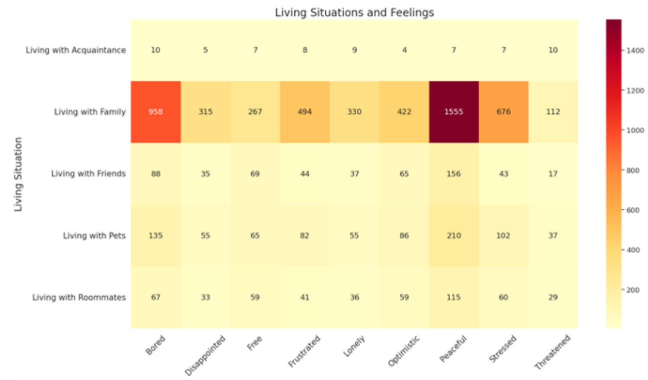

Figure 1: Heat Map for Different Living Situations and Feelings

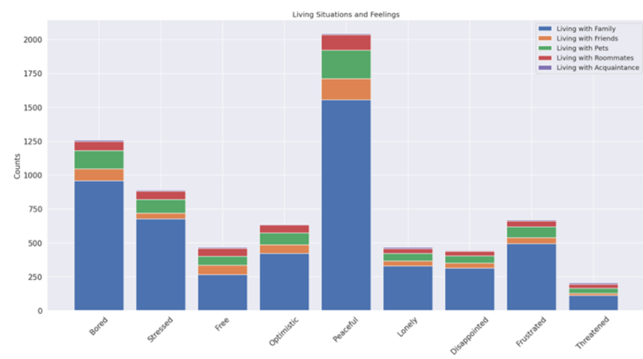

Figure 2: Stack Bar Graph for Different Living Situations and Feelings

4. FINDINGS IN THE MAIN DATASET

4.1 Emotional States in Different Living Situations During the Pandemic

To assess the emotional experiences and different living situations during the pandemic, respondents were asked, "During the lockdown, who do you live with?" The answers options where acquaintances, family, friends, pets, and roommates. A subsequent question, "How did spending time with them make you feel?" with responses options bored, disappointed, free, frustrated, lonely, optimistic, peaceful, stressed and threatened. The data were visualized using a heat map **Figure 1** and a stacked bar graph **Figure 2**, which shows the frequency of each emotion within the given living situations.

Analysis of **Figure 2** shows a higher incidence of peacefulness, boredom, and stress across living arrangements. Also, **Figure 1** demonstrates a consistent reporting of peacefulness irrespective of the living situation. The Mental Health Center corroborates these findings, suggesting that emotional support within family systems is pivotal for mental health well-being. Conversely Figure 2, indicates a substantial reporting of stress within familial environments. This could be reflective of the negative psychological impact due to unresolved familial conflicts, which are known to create hostile environments, chronic stress, and other mental health disorders, according to the Mental Health Center.[5]

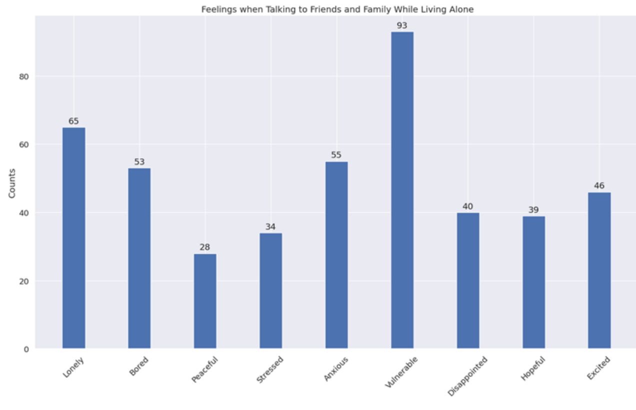

Figure 3: Bar Graph for Different Living Alone and Feelings

Additionally, individuals living alone were asked the question, "How did you feel when you talked to friends/family that you could not meet during the pandemic?" The responses, visualized in **Figure 3**, indicated that vulnerability, anxiety, and boredom were predominant. Psychology Today suggests that loneliness is not solely a factor of social connection quantity but also quality[10]. This may provide insight into the reported feelings of vulnerability and anxiety, which could be amplified by a negativity bias in social interaction perception among lonely individuals.

4.2 Agreement with COVID19 Prevention Activities

To capture public opinion on COVID19 prevention measures, respondents were surveyed regarding their level of agreement with various policies for 2020 and 2021. Specifically, they were asked, "How did you agree with the following rules in 2020 and 2021? (Mask mandate, Lockdown, Closing work/university, vaccination, social distancing, and quarantine)." Respondents could select a rank from 1 to 5, corresponding to 'Strongly Disagree' to 'Strongly Agree'. The resulting data were visualized through a heat map **Figure 4** and a horizontal stacked bar graph **Figure 5**, depicting the frequency of each level of agreement.

Analysis of **Figure 5** reveals a consensus in favor of most COVID19 prevention measures; however, a noticeable increase in strong disagreement in 2021 is seen, particularly with the policies involving the closure of workplaces and universities and lockdowns.

The heat map **Figure 4** contrasts the differences in response counts between the two years, providing the frequency of feelings. Higher counts of strong disagreement for workplace and university closures and lockdowns in 2021 are represented by lighter shades, suggesting a shift in public opinion. The heat map suggest a change in attitudes towards the extended preventive measures, with increasing levels of strong disagreement potentially reflecting economic pressures, mental health impacts, and a collective yearning for the resumption of normal activities after prolonged restrictions[15].

The survey asked participants "How did you agree with the following rules in 2020 and 2021? (Mask mandate, Lockdown, Closing work/university, vaccination, social distancing, and quarantine)." Answers were broken down by age, gender, and ethnicity, with respondents rating their agreement on a scale from 'Strongly Disagree' (1) to 'Strongly Agree' (5).

The results were then displayed in bar charts, categorized by these demographic factors.**Figure 6** shows the average levels of agreement, revealing demographic patterns in attitudes towards the pandemic measures. The data segmented by age indicates varying levels of consensus among different groups, with younger individuals (below 30) generally showing higher agreement. Gender-based analysis reveals that males and females, on average, show similar levels of agreement across all measures.

Table 1: Total positive count and percentage of people who tested positive for Covid 19 in the years 2020, 2021 and 2022

Year Tested Positive	Count	Percentage
Positive 20	371	22.84
Positive 21	443	27.28
Positive 22	810	49.88

4.3 Attitudes Towards Vaccination Status

In assessing public sentiment toward individuals who did not want to receive the COVID19 vaccine, participants were asked, "How did you feel about people not getting the vaccination?" **Figure 7** illustrates the bar graph for the emotional counts with 'Disappointed' being the predominant emotion, these emotions are selected by over 2000 people reflecting a widespread sense of dissatisfaction with people not wanting to get the vaccination. Figure 7 also shows a small demographic of participants selecting being proud. The bar graphs, Figure 7, also highlights the collective negative emotional response to vaccination hesitancy.

Further demographic analysis is presented in **Figure 8**, which segregates the emotional responses by age group and gender to show that most groups have about the same proportions for each feeling.

4.4 Perceptions of Test Positive for COVID19

4.4.1 Frequency on testing positive for COVID19

Let's investigate the data in Table 1, the total positive count and %age of people who tested positive for Covid 19 in 2020, 2021, and 2022, and visualization in **Figure 9**. The data highlights a substantial surge in positive cases, with the peak occurring in 2022, reaching 810 points. Notably, this accounts for nearly 50 % of the total cases within the student population as shown in **Figure 10 and 11**, collectively providing a snapshot of Positive COVID19 Tests within the student population. The analysis further reveals a pronounced escalation in positive cases, specifically among individuals aged 20-24. These findings underscore a significant increase in COVID19 infections, mainly in the young adult demographic. The prominence of positive cases in 2022 emphasizes the need for targeted interventions and public health measures within educational settings, focusing on the 20-24 age group.

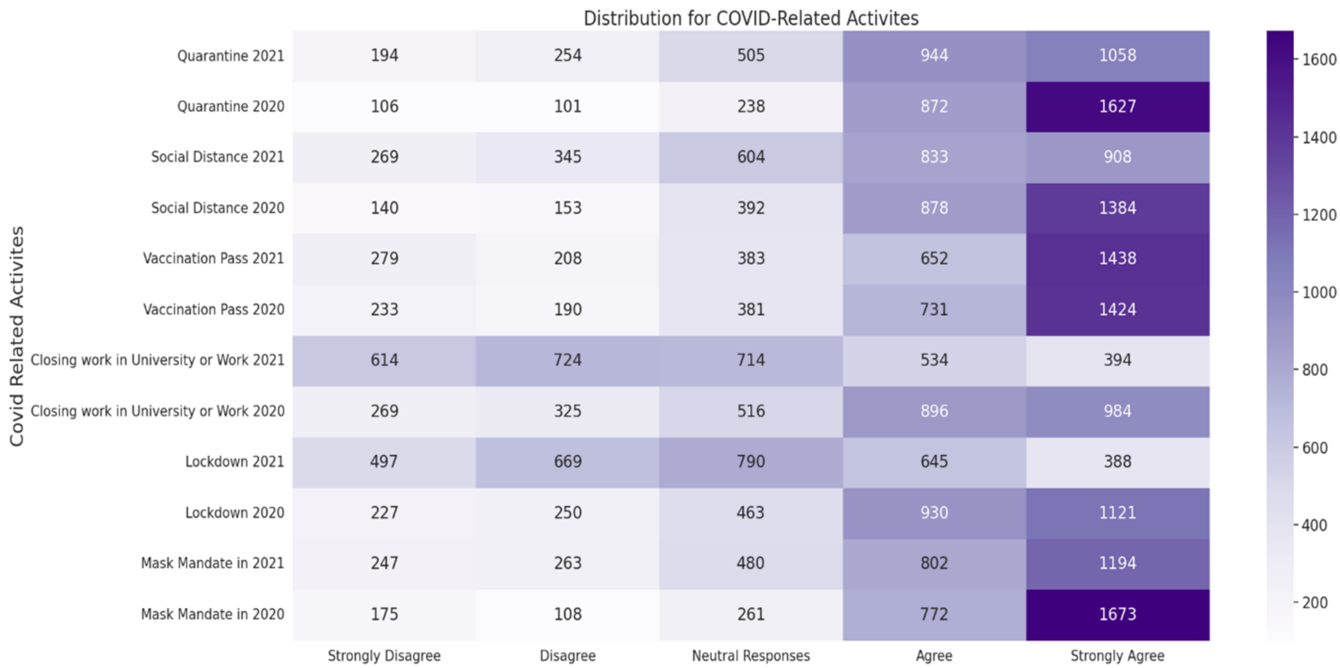

Figure 4: Heat Map for Agreement Levels and Covid Prevention Activities

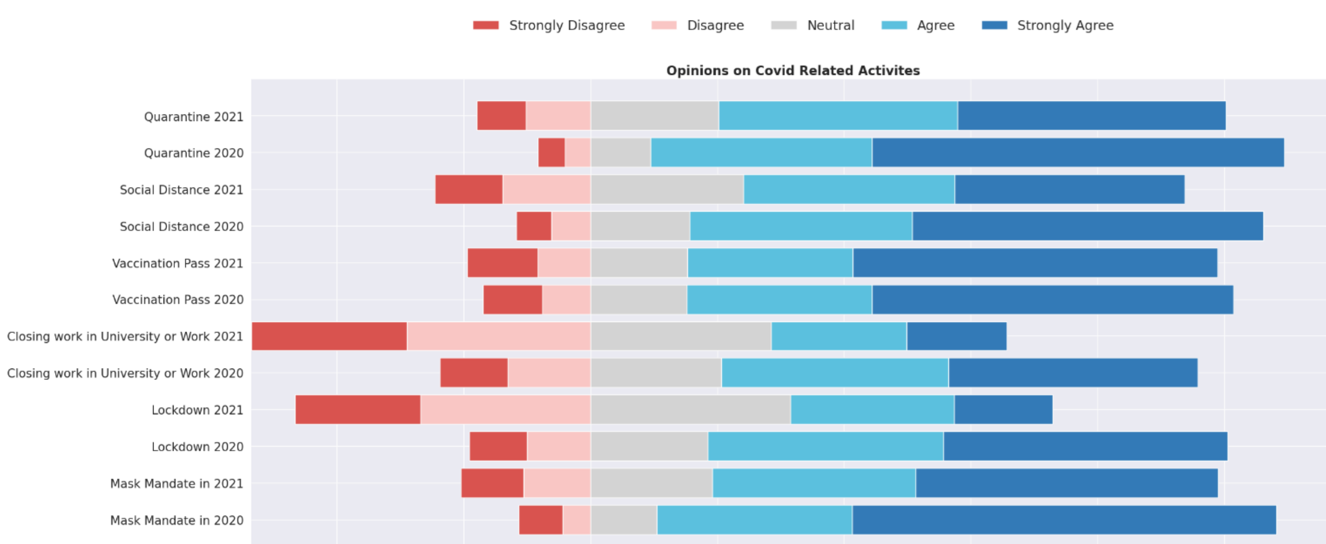

Figure 5: Horizontal Stack Bar Graph for Agreement Levels and Covid Prevention Activities

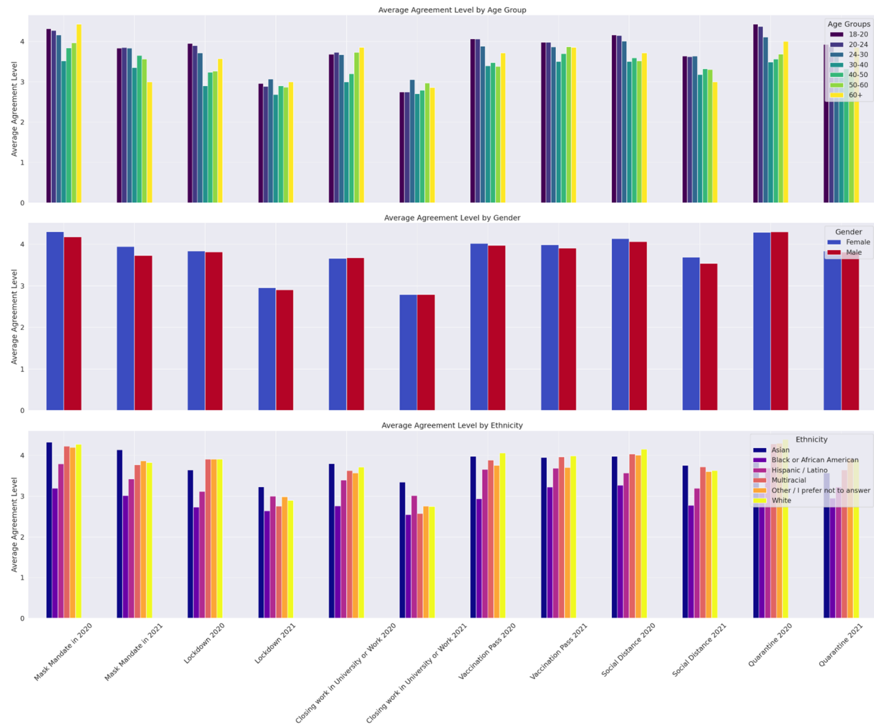

Figure 6: Bar Graph for Agreement Levels and Covid Prevention Activities by Age, Gender and Ethnicity

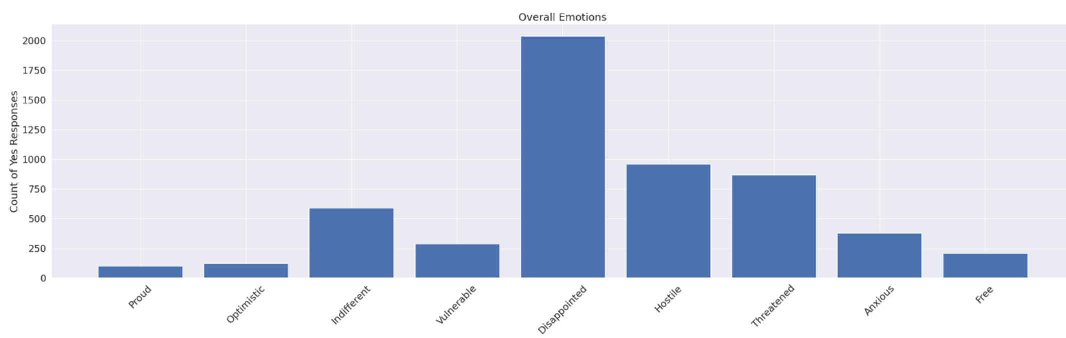

Figure 7: Bar Graph and Stack Bar Graph for Agreement Levels and Unwillingness to get Vaccinated

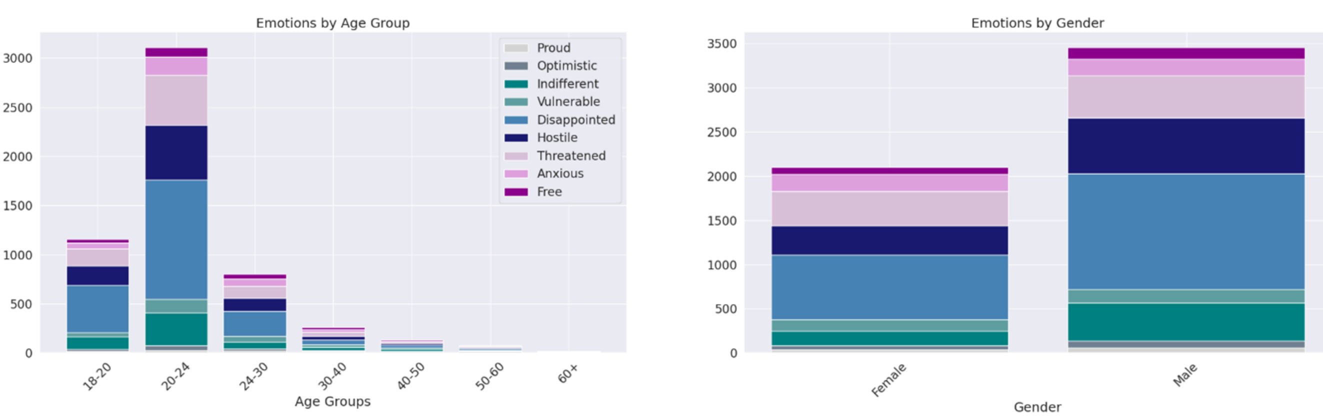

Figure 8: Stack Bar Graph for Agreement Levels and Unwillingness to get Vaccinated by Age and Gender

Table 2: Total count and the percentage of people who tested positive for Student and Professional by Year

Total times Tested Positive	Count	Percentage
Positive Once	1411	86.88
Positive Twice	167	10.28
Positive Thrice	46	2.83
Total	1624	100

Table 3: Total count and the percentage for different feelings of people who tested positive for COVID19

Emotion	Count	Percentage
Peace	546	33.62%
Frustrated	465	28.63%
Lonely	420	25.86%
Stressed	386	23.77%
Vulnerable	264	16.26%
Anxious	236	14.53%
Threatened	191	11.76%
Optimistic	187	11.51%
Disappointed	125	7.70%

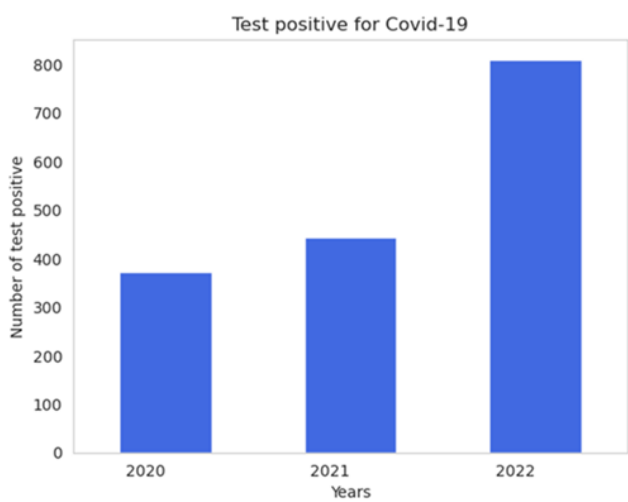

Figure 9: Bar graph for year and total count of positive test of COVID19

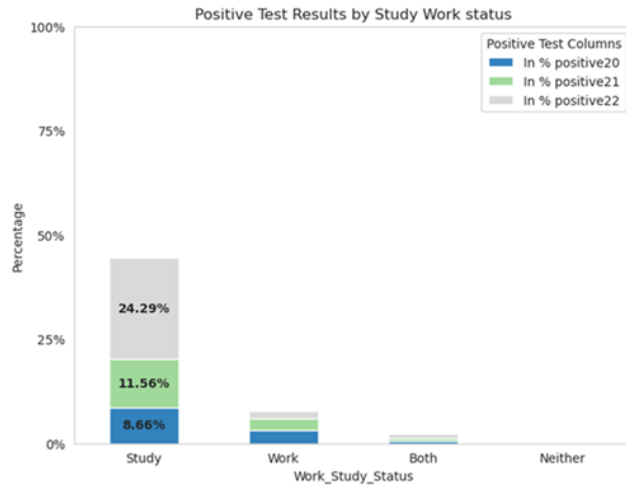

Figure 10: Stacked bar graph for Work or Study status of %age of positive COVID19 test by year.

Table 4: Total Count and the Percentage for Feelings Categorized as Positive, Negative, and Neutral Towards Studying of Students During Pandemic

Emotion	Count	Percentage
Positive	2308	77.76
Negative	356	11.99
Neutral	304	10.24

4.4.2 Emotional responses due to testing positive for COVID19

Additionally, the overview of Positive COVID19 Tests shows in **Figure 12** that 87 % had COVID19 once either in 2020, 2021, or 2022, while approximately 12 % had it twice. And these are related to people's emotional experiences when they become sick. The top 3 ranked feeling peaceful at 33.62 %, followed by expressing frustration at 28.63 % and loneliness at 25.86 %.

4.4.3 Emotional responses for students during COVID

We take a close look at how students feel about studying during the pandemic, as shown in **Figure 13**. The findings indicate that among students who had to study during this period, the most common emotions reported were Stress, followed by Frustration and Boredom. If we want to put all those emotions into the category of feeling, we begin with the starting point as Neutral and set it to a default category to ensure that every row has a starting point, It's like saying, "Let's begin by considering everyone in the middle, and then we'll figure out where they belong based on their actual feelings." Some feelings might not fit into positive or negative categories right away. This contributes to a Negative Categorization of their feelings, accounting for about 77 %, Neutral at 11.99 %, and Positive at 10.24 %.

4.4.4 Emotional responses for professionals in the work environment

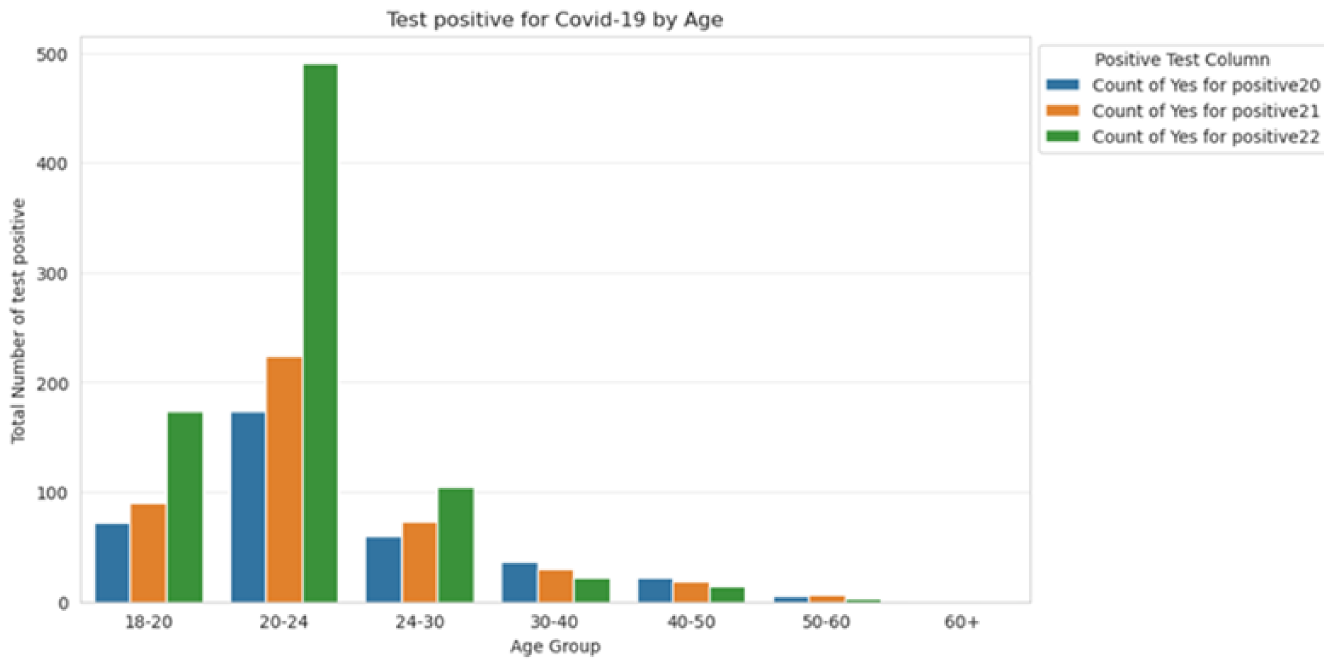

Figure 11: Bar Graph for Age Group to show the total count of positive COVID19 test by year.

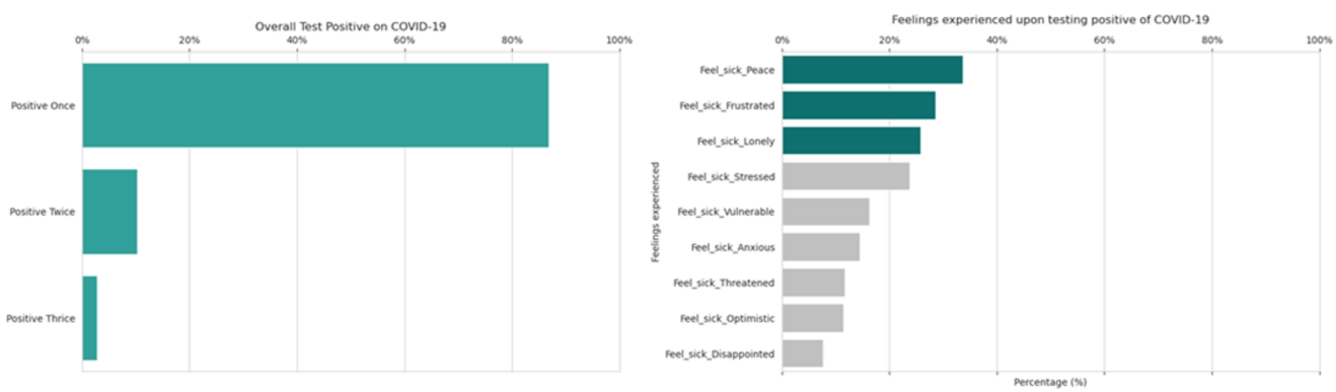

Figure 12: Horizontal Bar Graph for the %ages on Number of Times Tested Positive for Covid and Feelings Associated

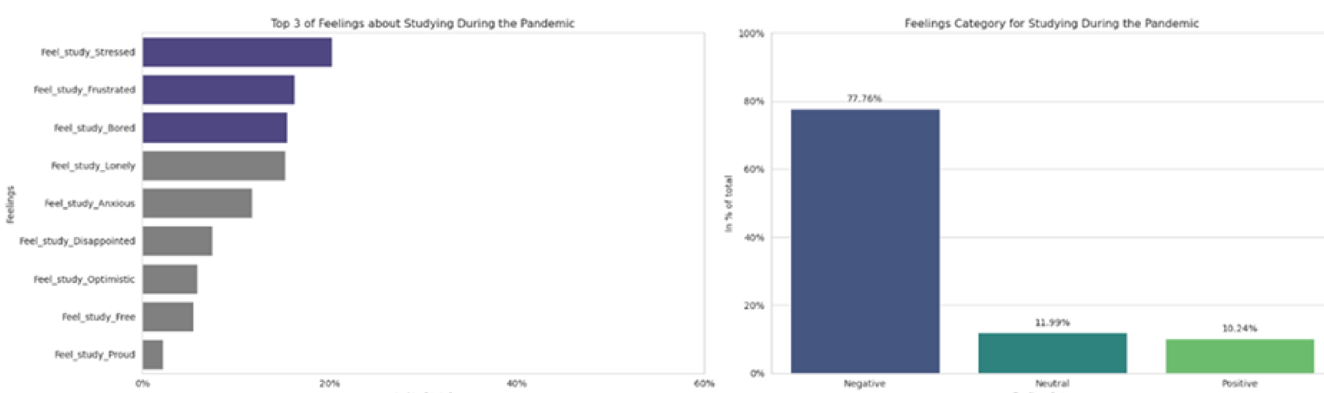

Figure 13: Horizontal Bar Graph for Different Feelings towards Studying of Students During the Pandemic and Bar Graph for Feelings Categorized as Positive, Negative, and Neutral

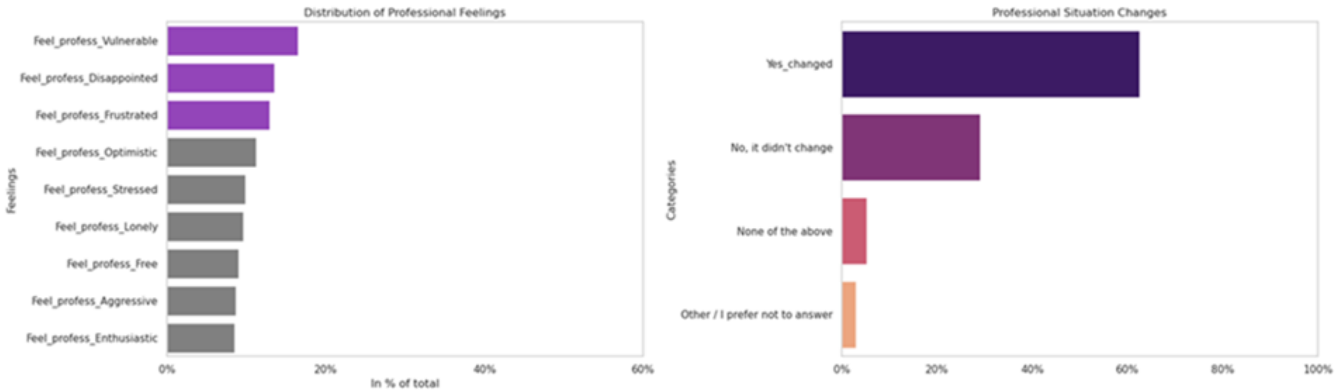

Figure 14: Horizontal Bar Graph for Different Feelings towards Professional Work Environment and Change of Professional Situation

Table 5: Count and Percentage of Feelings Towards Work Environment During Pandemic

Emotion	Count	Percentage
Vulnerable	109	16.49
Disappointed	88	13.31
Frustrated	86	13.01
Optimistic	74	11.20
Stressed	65	9.83
Lonely	64	9.68
Free	60	9.08
Aggressive	58	8.77
Enthusiastic	57	8.62

Table 6: Count and Percentage for Change of Professional Situation During Pandemic

Situation Change	Count	Percentage
Yes Changed	209	62.57
No, It Didn't Change	97	29.04
None of the Above	18	5.39
Other/I Prefer Not to Answer	10	2.99

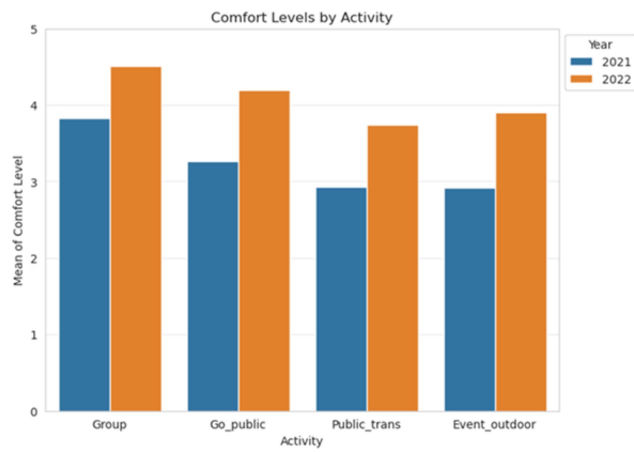

Figure 15: Bar Graph for Average Comfort Level for Different Activities During the Pandemic for Years 2021 and 2022

For the Emotional responses for professionals in work environment, as we move along, we will investigate the emotional responses of professionals in the workplace during the pandemic and examine the changes in their work situations. The results indicate that their emotions were Vulnerable, but disappointment and frustration were also common. The study reveals changes in their work situations, with over 62 % experiencing significant changes, such as job loss, quitting, reduced working hours, lower wages, smart working (remote work), or the unfortunate closure of their workplace, and about 29 % that did not change at all.

4.5 Comfort levels in social activities after lockdown COVID19

How comfortable do people feel with social activities in Winter 2021 and Summer 2022? In this part, the dataset was ranked on a scale of 1 to 5, where 1 means strongly disagree, and 5 means strongly agree. From Figure 16 our analysis reveals that people in 2022 are more willing to participate in social activities, especially in outdoor events, public transport, public spaces, and group gatherings, compared to 2021.

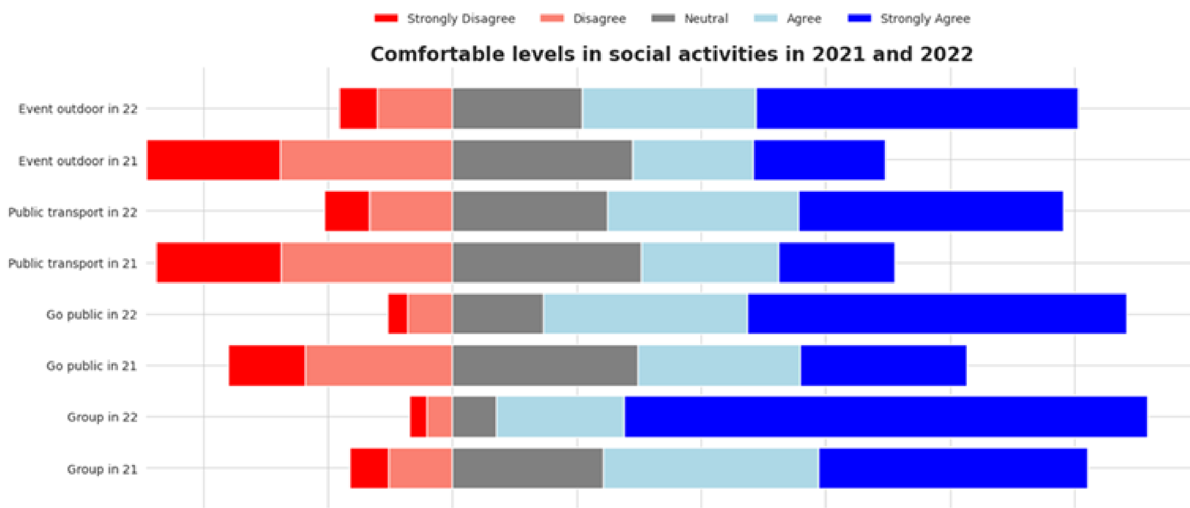

Figure 16: Stacked Bar Graph for Agreement for Different Activities During the Pandemic for Years 2021 and 2022

4.6 Perception of normality after Covid

Figure 17 presents a histogram depicting the distribution of individual responses on a numerical scale, assessing the perception of returning to one's pre-pandemic state of normalcy. The x-axis represents the scale of normalcy ranging from 1 ("Not normal at all") to 10 ("Completely normal"), and the y-axis quantifies the number of individuals corresponding to each point on the scale.

The data reveals a bimodal distribution, with the majority of individuals reporting a high level of normalcy, as indicated by the peaks around scores 8 and 10, which correspond to 30.8% and 20.8% of the respondents, respectively. There is also a notable minority reporting a neutral perception of normalcy (score 6), comprising 16.5% of the sample.

A significant observation from this histogram is the relative scarcity of individuals at the lower end of the scale, with those feeling not at all normal (score 1) representing merely 2% of the respondents. This could suggest that while a majority feel they have largely regained their pre-pandemic sense of normalcy, a portion of the population remains ambivalent.

This figure is essential for understanding the subjective sense of recovery in the post-pandemic period, and it can be instrumental in guiding further inquiry into the factors that contribute to this sense of normalcy or lack thereof. Additionally, it may inform targeted interventions to aid those who continue to feel detached from their pre-pandemic selves.

4.6.1 Comparative Analysis of Perceptual Normalcy Between Cities

The analysis reveals significant differences in the perception of normalcy post-pandemic between individuals in New York and Milan. As illustrated in **Figure 18**, the distribution of responses in Milan skews towards a higher sense of returning to normalcy compared to New York. This is substantiated by the positioning of the median lines on the box plots, which lean closer to the "Completely normal" end of the scale for Milan.

Factors Influencing Normalcy Perception by City. [3][5]

- Several factors may contribute to these observed differences:
- Severity and Duration of Lockdowns: New York's early and intense COVID19 outbreak necessitated longer and more stringent lockdowns, potentially causing a greater psychological and social impact.
- Government Response and Healthcare Systems: Variations in the pandemic response and healthcare system resilience between the cities may influence public sentiment.
- Cultural Differences: Local cultural attitudes towards the pandemic and the acceptance of new norms may affect adaptability.
- Economic Impact and Support: The economic toll and available support systems could shape the speed at which normalcy is perceived.
- Media and Communication: Media narratives and communication efficacy play roles in shaping the public's perception of the pandemic and its aftermath.
- Urban Environment and Lifestyle: The contrasting urban densities and lifestyles may impact the perceived return to pre-pandemic conditions.

Interpretation of Response Variability. The interquartile ranges and outliers observed in the box plots highlight the heterogeneity of individual experiences. The broader IQR for New York suggests a more diverse array of experiences with the pandemic's aftermath, while Milan's tighter IQR indicates a more uniform return to normalcy.

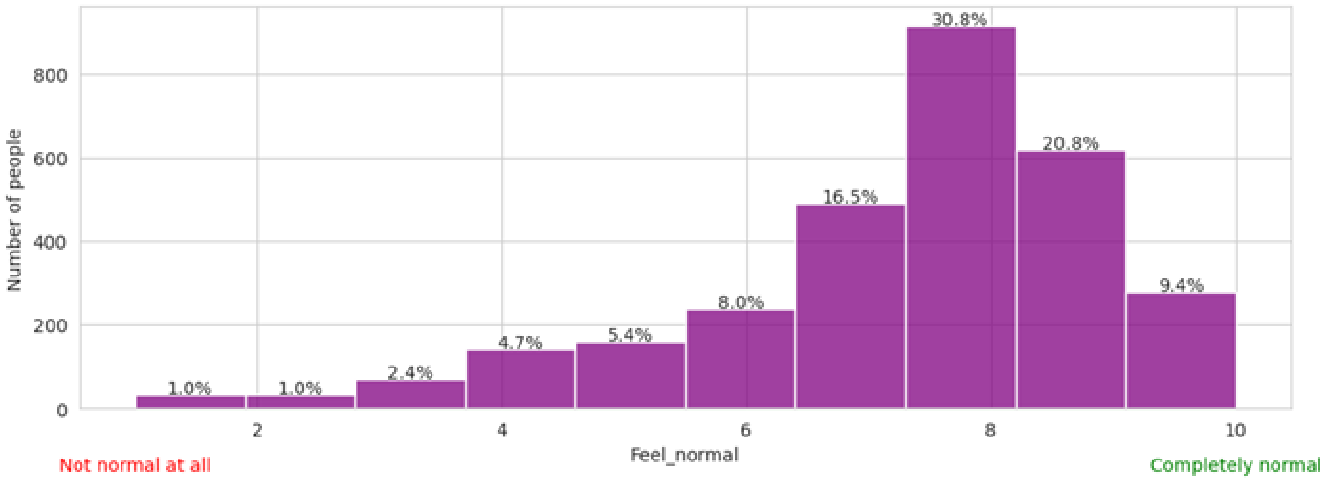

Figure 17: Histogram on a numeric scale the individual's perception of returning to one's pre-pandemic self

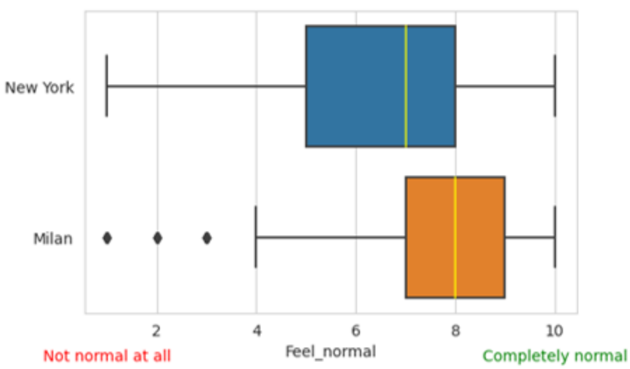

Figure 18: Box plot of perception of going back to normal by city

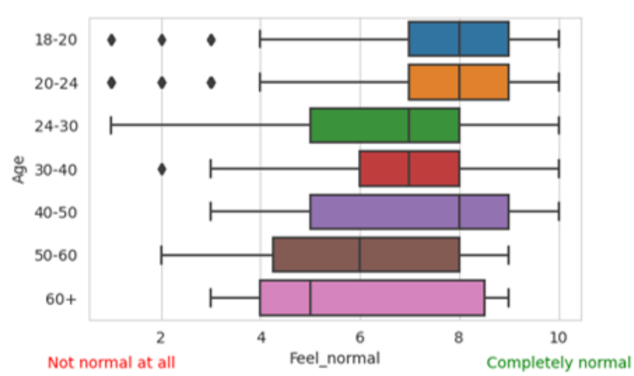

Figure 19: Perception of going back to normal after lockdown by age

4.6.2 Age-Based Dynamics of Normalcy Perception

The stratification of normalcy perception by age, as presented in the right panel of **Figure 19**, indicates that younger adults (18-24 years) exhibit a higher adaptability, with their responses tending towards higher normalcy scores. Conversely, individuals aged 60 and above show a marked decrease in normalcy perception.

Impact on the 60+ Demographic. The 60+ age group's lower sense of normalcy may be attributed to the direct and indirect impacts of the pandemic, such as increased mortality risk, health vulnerabilities, and the psychological ramifications of bereavement and isolation. These factors underscore the necessity for dedicated support and interventions for older adults, addressing the multifaceted challenges faced by this demographic in the wake of the pandemic.[1]

4.7 Predominant emotion that people living in New York city associate with COVID19

This section aims to elucidate the predominant emotional responses associated with COVID19 among residents of New York City. The analysis is based on a subset of a larger dataset that encompasses information from both Milan and New York, but herein we focus on the latter. Respondents were prompted to identify their emotions in relation to COVID19 from a given list.

4.7.1 Emotional Spectrum Analysis

Figure 20 is a horizontal bar chart that provides a detailed account of the range and frequency of emotions attributed to COVID19 by individuals in New York City. The x-axis quantifies the count of individuals who reported each emotion, while the y-axis lists the emotions, categorized as either negative or positive.

Observing the graph, it is immediately apparent that negative emotions dominate the emotional response to COVID19. 'Frustrated' is the most frequently reported emotion, followed closely by 'Anxious' and 'Stressed.' This suggests a significant level of emotional distress among the respondents, with these three emotions collectively encapsulating the predominant emotional state of the population.

On the lower end of the frequency spectrum, positive emo-

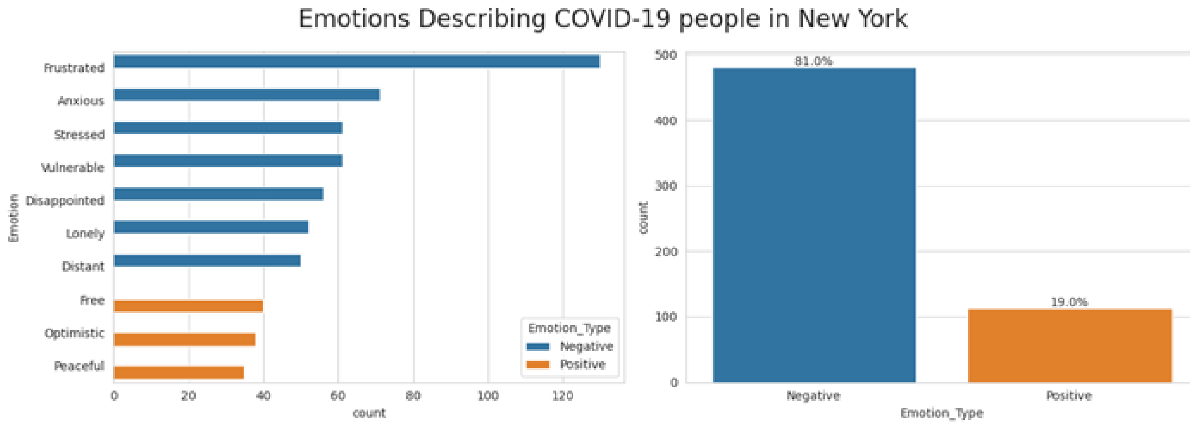

Figure 20: Range and frequency of emotions attributed to COVID19 by individuals in New York City

tions such as 'Free,' 'Optimistic,' and 'Peaceful' are present but notably less common. The presence of these emotions, although minimal, indicates that not all responses to the pandemic are negative and that a subset of the population retains a positive outlook despite the circumstances.

Provides the summary view, categorizing the emotional responses into negative and positive. It reveals a significant tilt towards negative emotions, with 81% of the recorded responses falling into this category, compared to 19% for positive emotions. This quantification not only confirms the predominance of a negative emotional response but also quantifies its magnitude, indicating a substantial impact on the mental health and emotional well-being of the city's inhabitants during the pandemic.

Job Sector-Specific Emotional Impact. **Figure 21** offers a breakdown of emotional responses by job sector. The data indicates a heterogeneous distribution of emotions across sectors, with information technology showing a higher frequency of negative emotions. This likely reflects the additional pressures faced by IT professionals during the pandemic, such as the sudden transition to remote work and the surge in demand for technological assistance and digital communication solutions.

Age-Related Emotional Trends. **Figure 22** explores the emotional responses stratified by age groups. Younger individuals exhibit a mix of negative and positive emotions, possibly suggesting better adaptability or resilience. In contrast, the positive emotional response diminishes progressively with increasing age, highlighting that older demographics may have been disproportionately impacted by the pandemic's challenges.

5. FINDINGS IN THE TWEETS DATASET

In this section, we delve into an advanced sentiment analysis of a Twitter dataset, contrasting it with our primary dataset which was geographically limited to New York and Milan. This comparative study leverages the enhanced capabilities of a customized ChatGPT plugin, specifically adapted for this purpose.

5.1 Creation of a GPT Assistant plugin to perform a sentiment analysis

- Objective: Development of a specialized data scientist assistant capable of performing sentiment analysis on the dataset.
- Iterative Enhancement: Progressive refinement of the model to increase accuracy and precision for the dataset, maintaining a global perspective.
- User Interaction Options: Introduction of four user-choice functionalities:
  - • Comprehensive sentiment analysis of the dataset.
  - • Analysis focusing solely on polarity and subjectivity.
  - • Analysis based on a predefined list of emotions.
  - • Generation of visual representations of sentiment data.
  - • A default response providing an overview of the assistant's functionalities is programmed for general inquiries (refer to attached **Figure 23**).

5.1.1 Selection of Analytical Analysis for the GPT Assistant

- Following a review of public Python libraries, we chose TextBlob for sentiment analysis. This decision was influenced at the same time by ChatGPT's proficiency in seamlessly integrating with TextBlob, offering a balanced approach between user-friendliness and analytical depth.
- Capabilities of the Customized GPT: Despite its specialized nature, this version of ChatGPT retains its inherent functionalities, including web browsing, DALL-E image generation, and code interpretation capabilities. This integration ensures a versatile and comprehensive analytical tool.

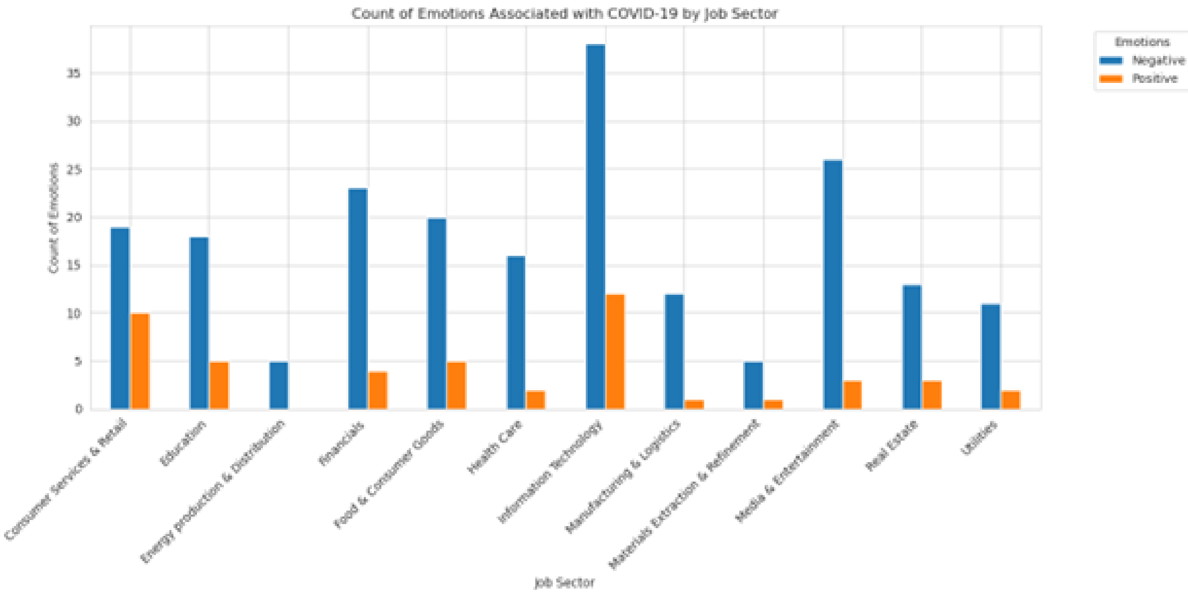

Figure 21: Emotional responses by job sector in New York city

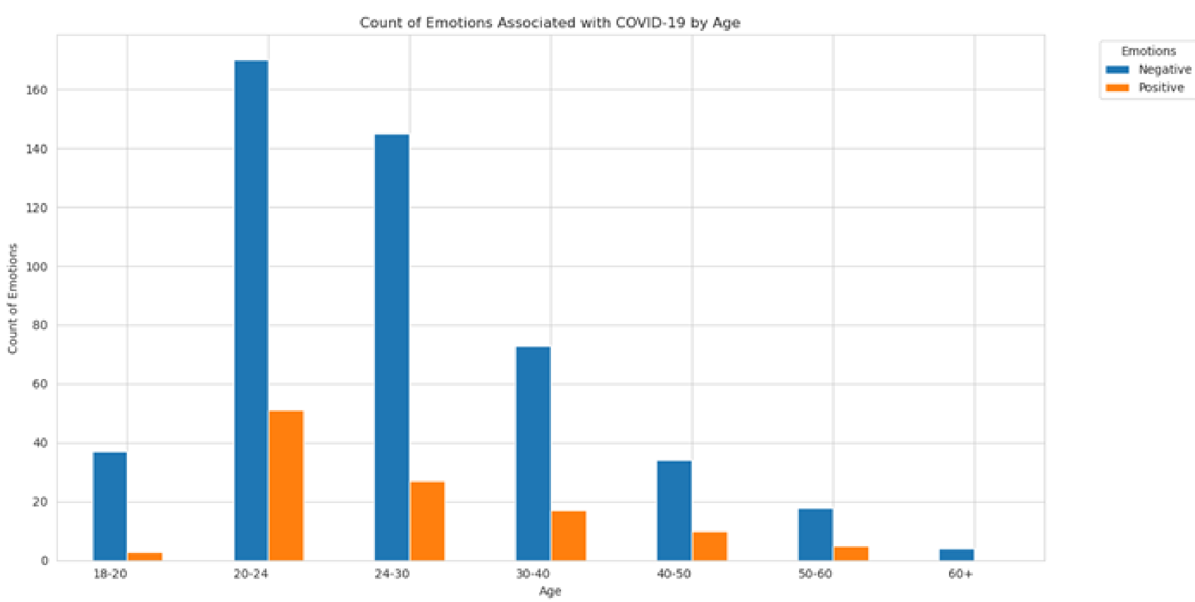

Figure 22: Emotional responses by age in New York City

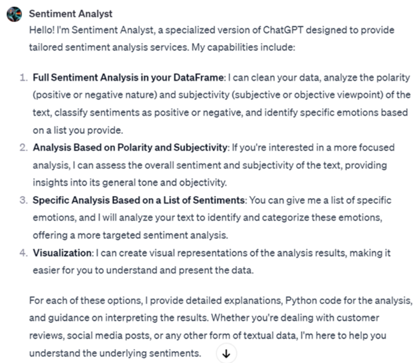

Figure 23: Sentiment Analyst GPT Assistant

5.2 Analysis Base on Polarity and subjectivity using TextBlob

This section focuses on the utilization of TextBlob[13], a Python library, to assess the polarity and subjectivity of tweets related to COVID19. Our methodology employs a specialized sentiment analysis function, analyze _ sentiment, which strategically bifurcates the dataset into different sentiment classes based on the content of the tweets.

The analyze _ sentiment function incorporates a regular expression search to identify any tweets that pertain to individuals "testing positive for COVID". These specific tweets are programmatically assigned a neutral polarity and subjectivity score. The rationale for this approach is the presumption that such tweets, while possibly containing emotional undertones, are primarily disseminating factual information.

For tweets that do not reference a COVID19 positive test result, TextBlob's sentiment analysis is applied. This analysis yields two key metrics: polarity, which ranges from -1 (most negative) to 1 (most positive), and subjectivity, which spans from 0 (objective) to 1 (subjective). Based on the polarity score, we categorize each tweet into one of three sentiment classes: positive, neutral, or negative.

The **Figure 24** derived from our analysis indicates a predominance of neutral sentiments, largely due to the initial classification rule. However, upon closer examination, particularly through a manual review of a sample of 100 tweets **Figure 24**, we find that the classifications of negative sentiments are notably more precise than those of positive sentiments. This discrepancy suggests that while TextBlob is adept at identifying negative sentiment cues, it may be less sensitive to the varied nuances of positive sentiment expression.

- For future analysis, a more granular approach might be necessary to distinguish between factual neutrality and emotional neutrality and to better capture the full spectrum of sentiments in tweets that contain both informational content and emotional expression.
- Additionally, to address the limitation in positive sentiment accuracy, it might be beneficial to refine the sentiment classification criteria or to incorporate supplementary linguistic models that can more effectively

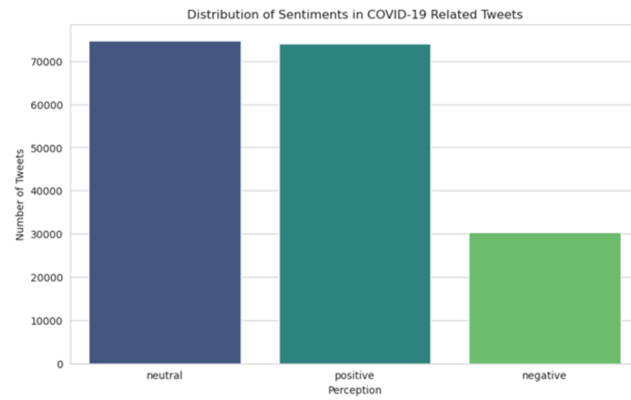

Figure 24: Sentiment Analysis Base on Polarity and Subjectivity using TextBlob in the Tweets Dataset

```
filter_rows_emotions(data_covid, 'perception', 'positive').head(20)
```

	text_cleaned	polarity	subjectivity	perception
1	hey and wouldnt it have made more sense to have the players pay their respects to the a	0.500000	0.500000	positive
6	how covid19 will change work in general and recruiting specifically via recruiting	0.050000	0.500000	positive
8	praying for good health and recovery of \ncovid19\ncovidpositive	0.700000	0.600000	positive
12	no one will be safe from covid19 until everyone is safe will you commit to ensure	0.500000	0.500000	positive
13	lets all protect ourselves from covid19\nits real and the numbers are climbing up fast in the continent\nlets n	0.200000	0.450000	positive
17	second wave of covid19 in flandersback to more homework again	0.250000	0.250000	positive
19	covid update the infection rate in florida is following the natural curve that experts predicted if the initial cu	0.033333	0.166667	positive
20	good patriots\ncall to volunteer to be an election judge \npolls cannot open without	0.350000	0.550000	positive

Figure 25: Small sample of tweets categorized as positive perceptions using TextBlob

interpret positive nuances.

5.3 Specific Analysis based on a list of sentiments

The primary goal of this analysis is to compare the emotions identified in this Twitter dataset with those in our main dataset. To achieve this, we employed three distinct methodologies: NLTK[8], WordNet[4], and manual analysis. These methods helped us identify specific emotions in our perception list from the previous dataset and compare the results across different approaches.

Each approach offers a unique perspective on sentiment analysis:

- Approach 1 leverages a sophisticated machine-learning model for nuanced sentiment analysis.
- Approach 2 broadens the scope by including synonyms and antonyms.
- Approach 3 offers a more controlled, variation-based categorization method.

5.3.1 Approach 1: TextBlob with NaiveBayesAnalyzer[12]

This approach utilizes the TextBlob library, integrated with the NaiveBayesAnalyzer from Natural Language Toolkit, specifically designed for sentiment analysis. It involves a two-step process: keyword based emotion categorization and sentiment classification.

Details:

- Initialization: The TextBlob object is initialized with NaiveBayesAnalyzer, which is pre-trained on movie reviews. This combination allows for a more nuanced analysis of sentiment.
- Function - categorize _ emotions:
  - Keyword Matching for Emotions: The function first checks the input text for specific emotion keywords. If found, these emotions are categorized and returned.
  - Sentiment Classification: If no specific emotion keyword is found, the function classifies the overall sentiment of the text as 'Positive', 'Negative', or 'Neutral', based on the NaiveBayesAnalyzer's analysis.
- Application: This function is applied to a COVID19-related dataset, specifically to a 'text _ cleaned' column, adding a new column with the emotion or sentiment classification in the column called 'emotions _ textblob _ nltk'

5.3.2 Approach 2: WordNet Synonym and Antonym Extension

This method expands the emotion categorization by using WordNet, a lexical database, to include synonyms and antonyms of emotion words, providing a broader range of emotion detection.

Details:

- WordNet Integration: The function get _ related _ words fetches synonyms and antonyms for given words from WordNet, enriching the emotion vocabulary.
- Extended Emotion Categorization:For each emotion in the original list, related words (synonyms and antonyms) are added to create an extended list.
- Emotion Detection: The function advanced _ emotion _ categorization scans the text for any of these extended list words, categorizing the text based on found emotions.
- Application: Similar to the first approach, this function is applied to the 'text _ cleaned' column of the COVID19 tweets dataset, and put the results in a column generated 'emotions _ wordnet'

5.3.3 Approach 3: Manual Variation-Based Categorization

This approach is based on manually defining common variations of emotion words, like different forms or related terms, and using these for emotion categorization.

Details:

- Emotion Variations: A dictionary emotion _ variations is created, mapping key emotions to their various forms and related terms.
- Simple Variation Emotion Categorization: The function simple _ variation _ emotion _ categorization searches for any of these variations in the text.
- Emotion Identification: If variations are found, the corresponding emotions are categorized.
- Application: This function, too, is applied to the 'text _ cleaned' column in the dataset, similar to the previous methods and puts the results in a column generated 'emotions _ manually'

6. CONTRASTING FINDINGS FROM THE TWO DATASETS

Figure 26 is a clustered bar chart comparing three different approaches to emotion categorization within a dataset of tweets. Each cluster of bars represents a different emotion, and each bar within a cluster corresponds to a different analytical approach.

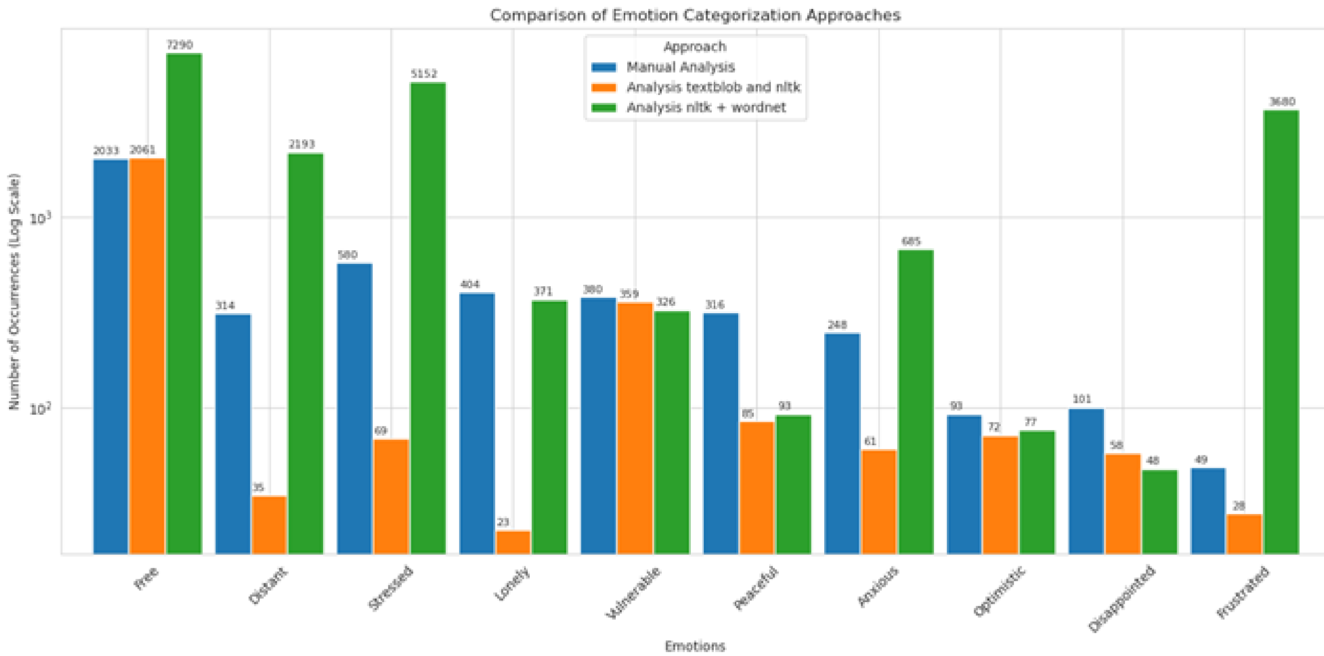

Figure 26: Perceptions and emotions identified in the tweets dataset

The y-axis is on a logarithmic scale, showing the number of occurrences (log scale) of each emotion as identified by each approach.

We can see the following:

- The emotion ”Free” has been identified most frequently by the Manual Analysis approach.
- ”Stressed” has a significantly high number of occurrences in the TextBlob and NLTK analysis.
- The NLTK + WordNet analysis has identified ”Frustrated” with the highest frequency among all emotions in its category.
- Manual Analysis seems to identify ”Free” and ”Frustrated” with much higher frequency compared to the other two approaches, which could suggest that the manual method is more sensitive to these emotions or that these emotions are more straightforward to identify without the nuances of linguistic analysis tools.
- The automated methods using TextBlob and NLTK, and NLTK with WordNet, show a more varied distribution across emotions like Lonely, Vulnerable, Peaceful, and Anxious.
- It’s worth noting that the emotion ”Optimistic” is identified relatively infrequently by all approaches, which could indicate that this emotion is either less expressed in the dataset or is more challenging to detect using these methods.

It’s noteworthy that while the first data source revealed negative perceptions, this second dataset predominantly shows positive perceptions. This discrepancy could be attributed to the limitations of our models, which tend to categorize tweets based solely on the presence of emotion-related words, potentially overlooking the actual context of the tweets. NLTK, TextBlob, and WordNet might have led to an overrepresentation of certain emotions. This was because the mere presence of words related to feelings in tweets was used to categorize them as expressions of those emotions, without a thorough contextual analysis.

7. CONCLUSIONS

- Sentiment Analysis of Tweets: The sentiment analysis of tweets indicates a trend toward positivity. However, this outcome might be influenced by the limitations of the dataset, which could be incomplete, and the potential lack of context in classifying emotions accurately.
- COVID19 and Mental Health: The analysis reveals that negative perceptions associated with COVID19 have a significant presence in the data. This negativity is correlated with an uptick in mental health issues.
- Resilience in Young Adults: The study finds that young adults, specifically those aged 20-24, have demonstrated resilience in the post-pandemic era. Despite facing a high rate of positive COVID19 tests, this demographic has maintained a semblance of normalcy.
- Employment and Stress During Pandemic: There is a discernible link between changes in employment due to the pandemic and heightened levels of vulnerability and stress. The disruption in employment status is strongly associated with these adverse emotional states.
- Benefits of Cohabitation: The data suggests that cohabitation during the periods of lockdown has been

more beneficial for mental health than living alone. This implies that the presence of cohabitants may provide emotional support and mitigate the stress associated with isolation.

8. FUTURE WORK

- Enhancement of Sentiment Classification: Future research will aim to develop a more sophisticated classification model that can more accurately distinguish between positive and negative sentiments, particularly in specific scenarios like the COVID19 pandemic. This will likely require the compilation and analysis of a larger and more diverse dataset.

- Improvements to GPT Bot: Work will continue refining the GPT-based bot to achieve higher accuracy in sentiment analysis. This refinement will include integrating more advanced Language Learning Models (LLMs) and potentially establishing direct connections to APIs. Utilizing state-of-the-art sentiment analysis tools, such as the 'twitter-roBERTa-base-sentiment' from Hugging Face, will be a key component of this effort.

- Data Expansion: There is a planned expansion of the current dataset to include a broader time frame and additional geographical locations. This expansion will encompass more years and cities, thereby enriching the dataset for more comprehensive analysis and improving the robustness of the findings.

9. REFERENCES

[1] Centers for Disease Control and Prevention. *COVID-19 Risks and Information for Older Adults*, 2023. https://www.cdc.gov/aging/covid19/index.html.

[2] L. Draisci, Y. Gao, F. F. Gonzales, B. Hu, X. Ma, E. Righini, H. Wang, M. Brambilla, S. Ceri, T. Davies, and M. Mauri. *Subjective Perceptions, Perspectives, and Feelings on the COVID-19 Pandemic in two US / EU Cities: Milan, Italy and New York City, USA.*, 2023. https://doi.org/10.7910/DVN/EWRL9K.

[3] J. Evelly. *Looking Back: COVID-19 in New York City, Two Years Later*, 2022.

[4] M. W. Goodman. *wn: A Python package for working with WordNet*, 2023. https://pypi.org/project/wn/.

[5] K. Hodge. *How Does Family Life Affect Mental Health? Building Blocks or Barriers*, Jun 2023. https://www.mentalhealthcenter.org/family-life-affect-mental-health/.

[6] Kaggle. *COVID19 Tweets*, July 2020. https://www.kaggle.com/datasets/gpreda/covid19-tweets/data.

[7] M. Nacoti, A. Ciocca, A. Giupponi, P. Brambillasca, F. Lussana, M. Pisano, G. Goisis, D. Bonacina, F. Fazzi, R. Naspro, L. Longhi, M. Cereda, and C. Montaguti. At the epicenter of the covid-19 pandemic and humanitarian crises in italy: Changing perspectives on preparation and mitigation. *NEJM Catalyst Innovations in Care Delivery*, Mar 2020.

[8] NLTK Project Contributors. *NLTK Documentation*, 2023. https://www.nltk.org/.

[9] OpenAI. *Creating a GPT — OpenAI Help Center*, 2023. https://help.openai.com/en/articles/8554397-creating-a-gpt.

[10] Psychology Today Staff. *Loneliness*. Reviewed by Psychology Today Staff.

[11] L. Sher. *The impact of the COVID-19 pandemic on suicide rates*. QJM : monthly journal of the Association of Physicians, April 2020. https://doi.org/10.1093/qjmed/hcaa202.

[12] TextBlob Contributors. *Advanced Usage: Overriding Models and the Blobber Class — TextBlob Documentation*, 2023. https://textblob.readthedocs.io/en/dev/advanced$_u sage.html.$

[13] TextBlob Contributors. *TextBlob: Simplified Text Processing*, 2023. https://textblob.readthedocs.io/en/dev/.

[14] C. N. Thompson, J. Baumgartner, C. Pichardo, and et al. Covid-19 outbreak — new york city, february 29–june 1, 2020. *MMWR Morb Mortal Wkly Rep*, 69:1725–1729, 2020.

[15] Wikipedia. *Pandemic fatigue*. https://en.wikipedia.org/wiki/Pandemic$_f atigue.$

[16] World Health Organization. *COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide*, Mar 2022. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence.