fortune Journals

Research Article

JOURNAL OF ENVIRONMENTAL SCIENCE AND PUBLIC HEALTH

ISSN: 2575-9612

Exploring U.S. Occupant Perception Toward Indoor Air Quality Via Social Media and NLP Analysis

Mehdi Ashayeri[1*], Soroush Piri[1], Narjes Abbasabadi[2]

Abstract

The global implementation of stay-at-home mandates altered people's activities within the built environment, prompting a slowdown in the spread of covid viruses. Nevertheless, this period shed light on previously unforeseen challenges in achieving "better" indoor air quality (IAQ) within buildings, necessitating a focus on building health resilience for future scenarios. This study aims to evaluate occupants' feedback on the impact of stay-at-home measures on IAQ perception in buildings across the U.S. during the first year of the pandemic (2020) and compare it with the baseline from the previous year (2019) nationwide to assess the changes and identify potential areas for IAQ management strategies. Geo-tagged textual data from X (formerly known as Twitter) platform were collected and analyzed using Natural Language Processing (NLP) based on time series sentiment analysis techniques to compute the feedback. Findings indicate that occupants' negative feedback on IAQ increased during 2020 compared to the baseline. It was also found that public perception of IAQ in 2020 was notably less favorable, potentially due to deteriorating conditions inside homes as people spent more time indoors. The study underscores the potential of NLP in capturing occupant perception, contributing to data-driven studies that can inform design, engineering, and policy-making for sustainable future.

Keywords: Indoor Air Quality; Occupant Perception; COVID Stay-at-home; Natural Language Processing (NLP); Time Series Sentiment Analysis.

Introduction

During the COVID-19 pandemic, the stay at home became a routine which helped prevent exposure to SARS-Cov2 viruses and mitigate the transmission rate [1]. According to the American Time Use Survey [2], the remote work during Covid-19 pandemic increased from 24% to 38% in 2020 compared with the baseline (2019). Due to the lockdown, in the U.S., people spent most of their time indoors: 90% in urban areas, 86% in the suburbs, and 82% in rural areas from March to April 2020 [3]. Accordingly, occupants started to complain on their buildings environments that they were not aware of them before the pandemic because the related challenge had not been raised in such scale [4]. Indoor air quality, that directly impact the health and well-being of building occupants [5], is among those building characteristics that occupants have largely expressed complaints on them during the lockdown. It measures the quality of indoor air in a building which affects the comfort and well-being of occupants. It is evident that exposure to indoor air pollution can negatively impact occupants' health (e.g., respiratory and cardiovascular diseases [6,7] [8], cognitive malfunction [9] and more).

Affiliation:

[1]School of Architecture, College of Arts and Media, Southern Illinois University, Carbondale, IL, United States.

[2]Department of Architecture, College of Built Environments, University of Washington, Seattle, WA, United States.

***Corresponding author:**

Mehdi Ashayeri, School of Architecture, College of Arts and Media, Southern Illinois University, Carbondale, IL, United States.

Citation: Mehdi Ashayeri, Soroush Piri, Narjes Abbasabadi. Exploring U.S. Occupant Perception Toward Indoor Air Quality Via Social Media and NLP Analysis. Journal of Environmental Science and Public Health. 8 (2024): 49-58.

Received: March 18, 2024
Accepted: March 26, 2024
Published: May 14, 2024

fortune Journals
Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

In addition, indoor air quality boosts virus transmission rates in buildings. Most recently, findings (e.g., [10]) indicate that the use of cleaning supplies in buildings along with the poor ventilation condition during the Covid-19 pandemic led to higher indoor pollution concentrations beyond the standard level [11]. The impacts of indoor pollution concentrations on occupants' health rose throughout the quarantine due to changes in outdoor concentrations level elevated before the pandemic and exposure to outdoor-origin indoor PM2.5 concentrations in buildings [12]. Among indoor pollutants, studies indicate that exposures to particulate matters (e.g., PM2.5) associated with higher mortality rates since PM2.5 particles can help carry the SARS-Cov2 viruses [13]. It further increases the severity of Covid-19 symptoms [14]. Therefore, indoor spaces needed to be properly ventilated, particularly with fresh air, to lower the virus spread rate and the risk of infection for future similar scenarios.

During the lockdown, dissatisfaction arose among certain groups due to inadequate Heating, Cooling, and Air Conditioning (HVAC) systems in their buildings. This was primarily because these buildings were not designed for prolonged stay-at-home scenarios [15]. Typically, HVAC systems are engineered to keep temperature, humidity, and contaminant levels within a standard range. However, adjusting these systems to accommodate extended use during the pandemic led to decreased efficiency and increased operation and maintenance costs [16]. One approach to mitigate this issue involves installing filters in the air-intake pathways, which can reduce the entry of outdoor particles into indoor spaces [17]. Additionally, high-efficiency particulate air (HEPA) purifiers can help maintain a clean air environment with acceptable HVAC performance [18]. However, their effectiveness against virus transmission is limited. This limitation stems from the fact that HEPA filters typically remove particles larger than 0.3µm in aerodynamic diameter, while SARS-Cov-2 virus particles are around or smaller than 0.1µm in size [19].

Moreover, studies confirm that there exists a direct correlation between higher temperature levels and indoor air quality dissatisfaction during the pandemic (e.g., [20,21]). An experimental study by [22] emphasize that a high temperature in indoor spaces increases volatile organic compounds (VOC) emissions, mostly driven by chemically-produced building materials, which result in poor indoor air quality, suggesting that occupant feedback on this topic occur more in the summer time, followed by them spring and winter. Negative feedback drops in winter due to the temperature decline [23, 24]. Study by [25] use data from the USGBC database to examine occupants' feedback on the indoor environmental quality (IEQ) of buildings. They compared buildings certified by the Leadership in Energy and Environmental Design (LEED) with non-certified buildings, employing text mining techniques for their analysis, highlighing that occupants' feedback, in terms of polarity scores (negative/positive), was nearly the same for both types of buildings.

Most recently, the collection of occupant feedback from online social data platforms has created new opportunities for research in indoor health. For example, research by [41] gathered data from online reviews of hotels on reservation websites and apply text mining techniques to capture occupant complaints on IEQ level in hotel rooms and find direct relationships between indoor air quality and seasonal weather variations as well as differences between indoor air quality satisfaction and climatic zones. And study by [26] extract big textual data from Airbnb reviews upon visitors' feedback on IEQ experiment during the their stay and apply text-mining approaches to measure the sentiment score. Similarly, study by [27] implement spatiotemporal text mining approaches to capture and evaluate occupants' feedback on IEQ dissatisfaction. In a recent study conducted by (Ashayeri and Abbasabadi 2024), occupant feedback concerning energy use in residential buildings was examined, with a focus on energy justice in New York City (NYC). Their work extracted data from X platform and utilized sentiment analysis.

Recent advancements in data-driven methods have facilitated the creation of more sophisticated models for human-centric design decision-making. Text mining, a key data-driven technique, leverages Natural Language Processing (NLP) frameworks. These frameworks enable machines to understand human language by transforming unstructured text into a structured format suitable for analysis [28]. Text mining offers timely and cost-effective data collection [29] and has been increasingly used in indoor health studies, especially in workplaces [30]. With the widespread availability of digital technologies and internet access in the past two decades, text mining of online reviews has gained popularity, providing access to a wider range of real-world data and deeper insights across various research domains [31,32,33].

While the utilization of social media as a source for big data is increasingly prevalent [34], and its role in influencing design decision-making and planning is well-recognized [35], there still remains a notable gap in leveraging such data for in-depth exploration of occupant feedback on indoor air quality during the Covid pandemic. This oversight underscores a deeper, intricate challenge in comprehensively grasping how prolonged indoor periods, typical of the pandemic era, have altered occupants' perceptions and experiences of air quality within their living spaces. Consequently, this research aims to delve into occupants' emotions regarding indoor air quality in buildings during the early Covid stay-at-home period. We developed NLP models based on sentiment analysis computational approaches to accurately capture the perceptions of building occupants towards indoor air quality at a time when their indoor presence was significantly higher than usual. This study utilizes geo-tagged Twitter data from

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

the first pandemic year (2020) and compares it with the baseline data from 2019, offering a deeper insight into the shifts in occupant experiences and perceptions during this unprecedented period.

Materials and Methods

The emergence of social media as a significant source of big data has transformed the landscape of research, design decision-making, and planning. Social media provides a unique platform that allows people from diverse background and geographical locations to share opinions, and experiences, making it an excellent resource for understanding human feedback on the built environment [36, 37]. Recent advances in data-driven techniques, such as NLP have opened up new opportunities for developing more reliable decision-making models. This paper applies big data and NLP approach to unravel public emotion on the indoor air quality in buildings during the early stages of the COVID-19 pandemic. This research used X platform, a major social media application that provides historical data for academic research [38]. The computational workflow for this research entails six major steps, 1) Data Extraction, 2) Data Cleaning, 3) Text Processing, 4) Emotion Analysis, 5) Performance Evaluation; and 6) Mapping and Visualizations:

Data Extraction

We selected X for social media analytics in this study due to several compelling reasons: 1) X users send short messages that are limited to 280 characters; therefore, people's emotion is conveyed through a short text communication; 2) Twitter offers a free product key to access historical data for the academic users [48], and 3) various studies have used Twitter as the primary social media platform for explorations of the built environment [49]. In this research, we used RStudio Programming Software for the computation. We used the *academictwitteR* library [50], as it allows for using the API v2 key –developed for the academic research– to extract historical data with minimum query restrictions. For performing the query, we highlighted the following terms to explore the indoor air quality: “air cleaner”, “air filter”, “air circulation”, “ceiling fan”, “co2 level”, “control air quality”,” dehumidifier”, “improve air quality”, “indoor air pollution”, “stale air”, “unhealthy air”, “virus airborne”, “iaq”, as well as “indoor air quality” itself.

Data Cleaning

After gathering the necessary data, we implemented a data cleaning protocol to render it suitable for application in the NLP simulation. This involved removing undesired characters such as URLs, signs, numbers, punctuations, and emojis, as well as English stop words such as "the", "is", "are", "he", "she", and other commonly used words. We utilized several R packages, including dplyr [42], tidyr [43], tm [44], and Corpus [45], to convert the users’ posts to cleaned texts. The goal was to ensure that the extracted data was free from noise and irrelevant information, thus increasing the accuracy of our subsequent analysis. The use of such packages for data preparation has been well-documented and widely adopted in previous studies on social media analytics [46,47].

Text Processing

The output of this step is a matrix known as the Document-Term Matrix (DTM), which is then passed to the subsequent step for NLP analysis. In emotion analysis, a DTM is a fundamental data representation used to analyze textual data, such as reviews, tweets, or any other form of text. It is a mathematical matrix that describes the frequency of terms (words or phrases) occurring in a collection of documents (texts). Each row of the DTM represents a document, and each column represents a unique term found in the entire corpus (collection of documents). The matrix cells contain the frequency count of each term in each document, indicating how many times a particular term appears in a specific document. Once the DTM is created, it becomes a valuable input for like emotion analysis techniques to classify emotions or emotions expressed in the social media comments. The DTM serves as a foundation for understanding the distribution of terms across documents and aids in identifying the emotion associated with those terms in the context of the text.

Emotion Analysis

In this study, we performed emotion analysis utilizing the Document-Term Matrix (DTM) developed in the preceding step, employing the SentimentAnalysis package available from the CRAN libraries. Emotion analysis and sentiment analysis are both subfields of NLP that deal with understanding human emotions and opinions from text. Sentiment Analysis is an application of NLP that analyzes language and extracts subjective information from narrative content to determine its relation to a decision variable [51]. This analysis can provide a continuous emotion score or a positive/negative classification. Sentiment analysis primarily aims to determine the polarity of a text—whether the expressed opinion is positive, negative, or neutral. Emotion analysis goes a step further by identifying specific emotional states expressed in a text. There are two primary methodologies for conducting sentiment analysis: machine learning-based approaches and dictionary-based approaches [52]. This work involves developing a dictionary-based model, which uses lexicons that contain words with predetermined polarities, chosen based on experts' intuition. This approach offers numerical scores that indicate the polarity and magnitude of the emotion. For emotion analysis, we used the *nrc* lexicon, which assigns words into ten emotion categories, including anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positive, and negative. It calculates the score for each category from sentences rather than just based on positivity or negativity scores. Table 1 lists the tribute of the *nrc* lexicon with the number of words per emotion category.

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

Performance Evaluation

This research involved evaluating the performance of emotion models by employing the Quantitative Discourse Analysis Dictionary (QDAP) [49] from the R's CRAN library. The QDAP package utilizes a comprehensive set of 1280 positive and 2952 negative words to effectively vectorize the (DTM) and facilitate accurate emotionanalysis. In our study, we employed three primary metrics to assess the performance of the emotion models, namely, Coefficient of Determination (R-Squared), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

The Coefficient of Determination (R-Squared) served as an essential metric to gauge the goodness- of-fit of the emotion analysis models. A higher R-Squared value indicated a better fit of the model to the emotion data, implying a more accurate representation of the variation in sentiment scores. The Root Mean Square Error (RMSE) was employed to measure the average magnitude of errors in predicting sentiment scores. Lower RMSE values were indicative of more accurate predictions, as they represented smaller average deviations between the predicted and actual sentiment scores. Similarly, the Mean Absolute Error (MAE) was utilized to calculate the average absolute differences between the predicted and actual sentiment scores. Lower MAE values indicated superior accuracy in emotion prediction, as they suggested smaller discrepancies between the predicted and actual emotion values. By incorporating these three comprehensive metrics and leveraging the powerful QDAP package, our study aimed to obtain robust and reliable insights into the emotion analysis process and assess the performance of the emotion models with precision and clarity."

Visualization and Mapping

In this research we performed various data visualization based on required analyses. We applied leaflet package for mapping geo-tagged tweets. We also used the RColorBrewer [50] and ggeasy [51] collectively package to visualize the results of word cloud analysis. We visualized results of the spatial SAs and word frequency using bar plots from the ggplot2 [44] package. And the scatter plot from the plotly package was used to depict the temporal SAs. All these R packages are free-accessed and can be downloaded through the Comprehensive R Archive Network (CRAN) database.

Results and Discussions

Figure 1 maps tweets related to indoor air quality-related keywords posted on Twitter pages within the U.S. boundaries during 2019 and 2020. As the maps illustrate, most of tweet instances occurred in large cities and their metro areas. Similarities on the spatial pattern between the two, indicatingthat the users' feedback are collected from almost the same locations.

This suggests a steady level of engagement with the topic of indoor air quality on X platform within these urban centers. Notably, the data indicates an uptick in the number of tweets from 2019 to 2020, particularly in large cities, perhaps signaling an increased public engagement or a response to events impacting air quality at the time. These maps, therefore, offer a visual summary of where and how discussions on indoor air quality are happening on Twitter, reflecting a persistent concern among inhabitants of more densely populated areas.

Figure 1 provides a geographical representation of tweets related to indoor air quality within the United States, captured during the years 2019 and 2020. The maps are marked with pink dots, each representing a tweet from a specific location, revealing a higher concentration of activity in urban areas. The continuity in spatial patterns from one year to the next suggests that the same metropolitan regions consistently contributed to the conversation on indoor air quality, indicating that users' feedback on indoor air quality was predominantly from the same cities across both years.

The bar chart in Figure 2-a quantitatively represents the frequency of tweets regarding indoor air quality on a monthly basis, revealing a higher incidence during the summer of 2019 as compared to the summer of 2020. The increase in 2019 is consistent with expectations that warmer conditions would amplify concerns related to indoor air quality, likely due to enhanced indoor discomfort and the resultant elevated usage of air conditioning systems related energy burden. However, the summer of 2020 exhibits a decline in such tweet frequencies despite the ongoing COVID-19pandemic, which would have presumably escalated attention to air quality matters due to the potential for airborne transmission of the virus. This decline suggests a shift in public attention, possibly prioritizing immediate pandemic-related issues over thermal comfort, thereby diminishing the prominence of conventicle perception of indoor air quality in public discourse. This deviation calls for a comprehensive investigation into the dynamics that dictate public engagement and the narrative surrounding indoor air quality amid concurrent public health emergencies.

Figure 2(b,c) word clouds represent the most frequently occurring words in the collected tweets. In both years, terms like "air," "conditioner," "ventilation," "system," and

Table 1: List of *nrc*-emotion categories with the corresponded number of sensitive words per emotion category

	nrc emotion categories									
Categories	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Positive	Negative
# words	1247	839	1058	1476	689	1191	534	1231	2312	3324

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

"HVAC" are prominent, which indicates these are key aspects of indoor air quality being discussed. The word "new" in the 2019 word cloud may suggest a focus on new installations or products. Meanwhile, in 2020, words like "covid" and "mask" emerge, highlighting the impact of the COVID-19 pandemic on public concern about indoor air quality and the inclusion of health safety measures in the discussion. The data from the bar chart combined with the word clouds suggest a heightened awareness and concern for indoor air quality, possibly intensified by the events of 2020, particularly the pandemic. The prominence of HVAC-related terms across both years signifies a consistent focus on the systems that manage indoor air. The integration of pandemic-related terms in 2020 underscores the influence of external health events on public discourse and concerns related to indoor environments. This analysis can provide insights into public interest trends and could inform public health messaging and policy-making related to indoor environmental quality.

Figure 3-a displays an aggregated sentiment analysis (SA) of tweets related to indoor air quality across the United States, conducted for the years 2019 and 2020. With a total of 22,798 tweets captured in 2019 and 22,505 in 2020, the bar chart reveals an uptick in negative sentiment scores in 2020 compared to the previous year. This suggests that public perception of indoor air quality during the pandemic lockdown was notably less favorable, potentially due to deteriorating conditions inside homes as people spent more time indoors, as indicated by the study from [52]. Figures 3-b and 3-c extend the analysis to monthly emotional responses using the NRC lexicon, corresponding with the tweet frequency data presented in a previous figure. The time series emotion analysis highlights a pronounced increase in negative sentiment in September 2020. This surge could be linked to various factors such as increased indoor emissions from more people working or studying at home, as suggested by [52]. Alternatively, this period's negative sentiment might be influenced by environmental factors like seasonal changes or external events like wildfires, which are known to affect air quality. This comprehensive SA underscores the importance of considering both temporal patterns and external events when evaluating public sentiment towards indoor air quality. The data indicates a complex interplay between various factors that impact public perception, with the global health crisis notably influencing sentiment in 2020. Further investigation into specific events or changes during this period could provide a more detailed understanding of the causes behind the observed shifts in sentiment.

Figure 1: Indoor air quality -related tweets and their distribution in the U.S. mainland in 2019 (left) and 2020 (right).

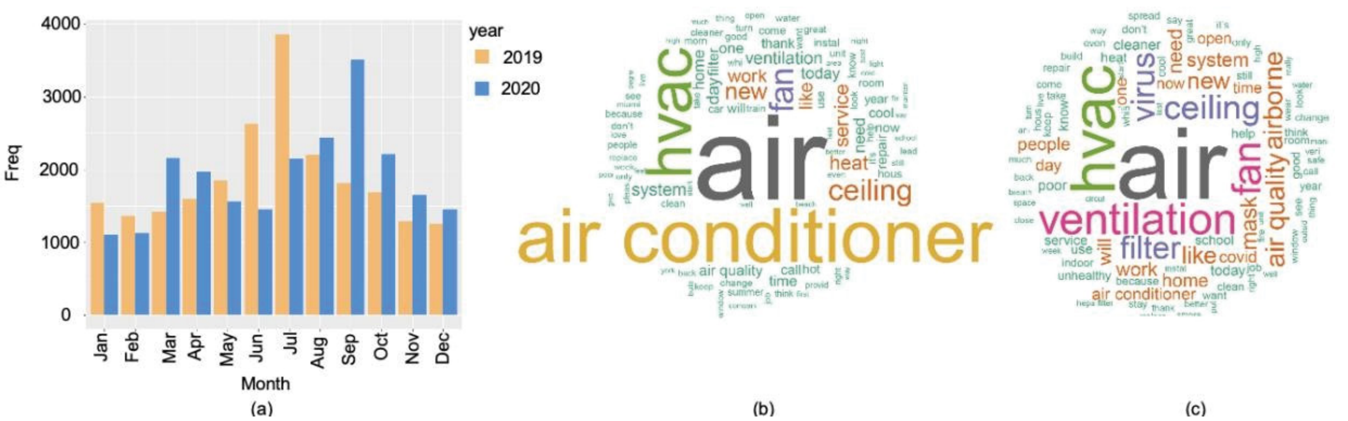

Figure 2: (a) Keyword frequencies upon cluster keywords explored on indoor air quality for 2020 vs. 2019. Wordcloud plots on cluster keywords queried for indoor air quality for 2019 (b) and 2020 (c).

Citation: Mehdi Ashayeri, Soroush Piri, Narjes Abbasabadi. Exploring U.S. Occupant Perception Toward Indoor Air Quality Via Social Media and NLP Analysis. Journal of Environmental Science and Public Health. 8 (2024): 49-58.

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

Table 2 presents the annual performance evaluations for 2020 and 2019, employing metrics such as R-squared, RMSE, and MAE. These evaluations were based on emotion analysis data, with a specific focus on negativity and positivity emotions, processed using the QDAP package. An examination of the R-squared values reveals that the model accounts for approximately 64.65% of emotion variation in 2020 and 67.14% in 2019. This level of explanation indicates a reasonably good fit to the data, albeit moderate. Upon a deeper inspection of the metrics, it becomes evident that the RMSE and MAE values for Negative Emotion in both years are comparatively higher, hinting at potential challenges in accurately predicting negative emotions. On the other hand, the model exhibits a more robust performance in predicting Positive Sentiment scores, as evidenced by the lower RMSE and MAE values relative to Negative Emotion. A comparison between the two years reveals a slightly better R-squared fit for 2019 at 67.14%, in contrast to the 64.65% in 2020. However, this minor discrepancy between the two years suggests that the model's performance has remained relatively consistent across the evaluated period.

Table 3 represents the performance metrics of temporal emotion analysis on Twitter data for the years 2019 and 2020, utilizing both the QDAP (Quantitative Discourse Analysis Package) and NRC lexicon. The QDAP, designed for quantitative discourse analysis, facilitates the transformation and analysis of textual data into numerical formats. While it is not primarily a sentiment analysis tool, it can be adapted and used in tandem with other tools or dictionaries, like the NRC lexicon, to evaluate emotions. The table details performance across three crucial accuracy metrics— R-squared, RMSE, and MAE—for overall emotion, negativity, and positivity for each month. In 2020, R-squared values for the overall emotion ranged from 0.6057 in February to a peak of 0.7128 in September, indicating the proportion of emotion variance captured by the model. Simultaneously, the RMSE and MAE metrics offer insights into the deviation between predicted and actual sentiment scores. The 2019 data exhibits a similar trend, with R-squared values shifting between 0.5997 in April to 0.6799 in February. This table provides a comprehensive perspective, enabling readers to evaluate the efficacy of the emotion analysis methodology used, leveraging both QDAP and the NRC lexicon, and to discern temporal performance patterns across the two years [53-65].

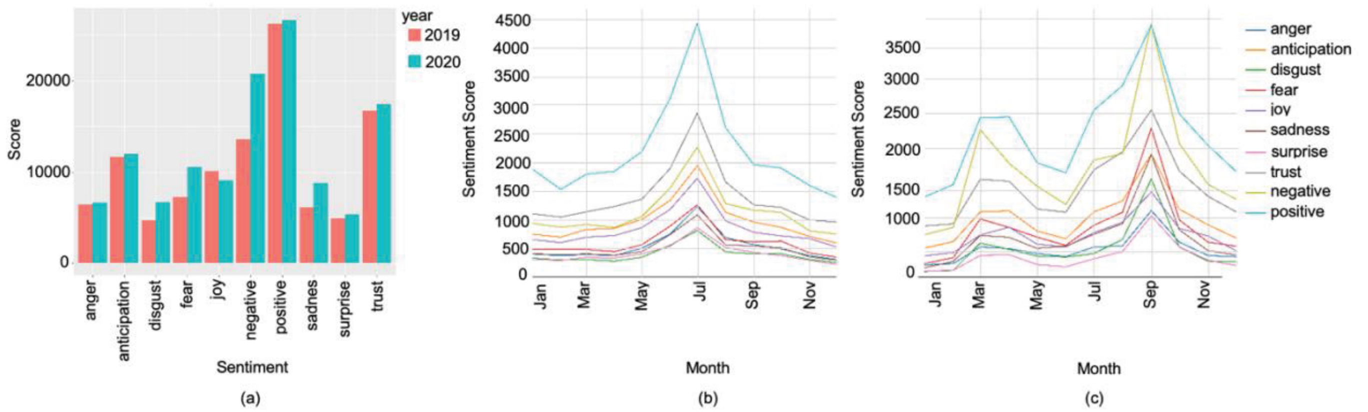

Figure 3: Sentiment Analysis of Tweets on Indoor Air Quality in the U.S. for 2019 and 2020. Panel (a) shows annual sentiment scores, with an increased negative sentiment in 2020. Panels (b) and (c) present monthly sentiment trends using the NRC lexicon, with a peak in negative sentiment observed in September 2020.

Table 2: Results of performance analysis on the R-squared, RMSE, and MAE metrics in 2020 and 2019.

Year	Performance Metric	Overall Emotion	Negative Emotion	Positive Emotion
2020	R-squared	0.6465	0.3842	0.3318
	RMSE	0.7651	0.8938	0.7798
	MAE	0.6694	0.793	0.6902
2019	R-squared	0.6714	0.4309	0.3421
	RMSE	0.7793	0.9152	0.8029
	MAE	0.6916	0.8243	0.7216

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

Table 3: Results of performance metrics for temporal emotion analysis based on QDAP dictionary and *nrc* lexicon for 2019 and 2020.

Month	Accuracy Metrics								
	R-squared			**RMSE**			**MAE**		
	Emotion	Negativity	Positivity	**Emotion**	Negativity	Positivity	**Emotion**	Negativity	Positivity
2020									
January	**0.65**	0.37	0.31	**0.75**	0.88	0.76	**0.64**	0.77	0.67
February	**0.61**	0.41	0.28	**0.78**	0.92	0.80	**0.70**	0.84	0.72
March	**0.69**	0.46	0.37	**0.77**	0.93	0.81	**0.69**	0.84	0.73
April	**0.66**	0.41	0.35	**0.78**	0.92	0.80	**0.71**	0.85	0.73
May	**0.65**	0.44	0.32	**0.77**	0.92	0.79	**0.69**	0.83	0.71
June	**0.66**	0.42	0.32	**0.77**	0.92	0.79	**0.68**	0.82	0.72
July	**0.66**	0.42	0.34	**0.78**	0.91	0.80	**0.70**	0.83	0.72
August	**0.65**	0.42	0.36	**0.77**	0.92	0.79	**0.69**	0.83	0.73
September	**0.71**	0.45	0.38	**0.77**	0.92	0.81	**0.68**	0.83	0.72
October	**0.68**	0.45	0.33	**0.77**	0.92	0.80	**0.68**	0.82	0.72
November	**0.65**	0.40	0.32	**0.77**	0.91	0.79	**0.68**	0.81	0.70
December	**0.67**	0.43	0.33	**0.77**	0.91	0.79	**0.68**	0.82	0.71
2019									
January	**0.61**	0.40	0.29	**0.76**	0.90	0.78	**0.67**	0.80	0.69
February	**0.68**	0.40	0.36	**0.76**	0.90	0.78	**0.66**	0.80	0.69
March	**0.62**	0.38	0.30	**0.77**	0.90	0.78	**0.67**	0.80	0.69
April	**0.60**	0.37	0.31	**0.77**	0.90	0.78	**0.68**	0.81	0.70
May	**0.63**	0.38	0.31	**0.76**	0.89	0.77	**0.66**	0.79	0.68
June	**0.65**	0.41	0.34	**0.76**	0.90	0.78	**0.67**	0.81	0.69
July	**0.66**	0.40	0.37	**0.75**	0.89	0.77	**0.66**	0.79	0.68
August	**0.65**	0.38	0.36	**0.75**	0.90	0.77	**0.66**	0.80	0.68
September	**0.65**	0.43	0.33	**0.76**	0.90	0.78	**0.66**	0.81	0.69
October	**0.65**	0.41	0.34	**0.76**	0.90	0.77	**0.66**	0.80	0.69
November	**0.65**	0.35	0.32	**0.76**	0.89	0.77	**0.67**	0.80	0.69
December	**0.66**	0.35	0.33	**0.74**	0.88	0.76	**0.64**	0.76	0.66

fortune Journals
Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

Limitation

Nevertheless, it is essential to recognize the general limitations inherent in emotion analysis studies. This method may not capture the full depth of human emotions, and the analysis of tweets might not provide a comprehensive insight into individuals' perceptions. Furthermore, it is worth noting that the data utilized in this study were not pre-screened based on specific building types. This decision was based on the assumption that a substantial portion of the population remained at home due to lockdown measures. However, this assumption may introduce variability in the dataset and fail to account for differences in indoor air quality experiences across various building types. Consequently, for a more rigorous exploration of the relationship between socio-economic factors and spatiotemporal emotion trends, further research is warranted.

Conclusion

This study examines the occupant feedback toward indoor air quality during the prolonged stay-at-home periods. Through sentiment analysis, the research brought to light a notable upsurge in dissatisfaction during the early stages of the pandemic compared to the preceding year. Furthermore, the study revealed that public perception of indoor air quality in 2020 experienced a significant decline in favorability, likely attributable to worsening conditions within buildings as individuals increasingly remained indoors. Architects and engineers can explore innovative ventilation systems and cost-effective solutions that promote better air quality for prolonged building operational use. Policymakers and planners should consider implementing regulations and guidelines that prioritize indoor air quality standards in building codes and planning initiatives. Additionally, raising public awareness about the importance of indoor air quality and providing resources for improving building ventilation and filtration systems can empower individuals to take proactive measures to safeguard their health. By collectively addressing these challenges and adopting such measures, initiatives can strive towards creating healthier and more resilient built environments for sustainable future. The study further emphasizes the potential of NLP and social media data in capturing human feedback, emphasizing the significance of data-driven linguistic analysis in the field.

Acknowledgements

For this research, Mehdi Ashayeri received partial financial support from the School of Architecture and the Office of Vice Provost and Vice Chancellor for Academic Affairs at Southern Illinois University Carbondale (SIUC). Narjes Abbasabadi received partial financial support from the Department of Architecture at the University of Washington (UW). Mehdi Ashayeri contributed to conceptualization, methodology, literature review, software, formal analysis, investigation, resources, data curation, original draft writing, visualization, and project administration. Soroush Piri served as research assistant, contributing to software, visualization, and literature (partially). Narjes Abbasabadi contributed to methodology, resources, investigation, and writing (review and editing).

References

1. Navas-Martín MA, Oteiza I, Cuerdo-Vilches T. Dwelling in Times of COVID-19: An Analysis on Habitability and Environmental Factors of Spanish Housing. J Building Engineering 60 (2022): 105012.
2. American Time Use Survey - 2021 Results 26 (2022).
3. Schaul K, Brittany RM, Bonnie B. Where Americans Are Still Staying at Home the Most (2022).
4. Kakoulli C, Alexis K, Michaelides MP. A Review of Field Measurement Studies on Thermal Comfort, Indoor Air Quality and Virus Risk. Atmosphere 13 (2022): 191.
5. World Health Organization. Indoor Air Pollutants: Exposure and Health Effects. EURO Reports and Studies 78. Copenhagen (1982).
6. Brook RD, Rajagopalan S, Pope CA, et al. Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. Circulation 121 (2021): 2331–2378.
7. Pope CA, Richard TB, George DT, et al. Cardiovascular Mortality and Long-Term Exposure to Particulat Air Pollution: Epidemiological Evidence of General Pathophysiological Pathways of Disease. Circulation 109 (2011): 71–77.
8. Pope CA, Nathan C, Zachari AP, et al. Corrigendum to 'Fine Particulate Air Pollution and Human Mortality: 25+ Years of Cohort Studies. Environ Res 183 (2020): 108924.
9. US EPA. Improving Indoor Air Quality. Overviews and Factsheets. https://www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality (2014).
10. Allen, JG. MacNaughton P, Satish U, et al. Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments. Environ Health Perspect 124 (2016): 805–812.
11. Mohamad A, Becerik-Gerber B, Hoque S, et al. Ten Questions Concerning Occupant Health in Buildings during Normal Operations and Extreme Events Including the COVID-19 Pandemic. Building Environ 188 (2021): 107480.

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

12. Bidila T, Pietraru RN, Ionita AD, et al. Monitor Indoor Air Quality to Assess the Risk of COVID-19 Transmission. CSCS 10 (2021): 356–361.

13. Li J, Yatai M, Xinlei L, et al. Field-Based Evidence of Changes in Household PM 2.5 and Exposure during the 2020 National Quarantine in China. Environ Res Lett 16 (2021): 094020.

14. Wu X, Nethery RC, Sabath MB, et al. Air Pollution and COVID-19 Mortality in the United States: Strengths and Limitations of an Ecological Regression Analysis. Sci Adv 6 (2020): eabd4049.

15. Benmarhnia T. Linkages Between Air Pollution and the Health Burden From COVID-19: Methodological Challenges and Opportunities. Am J Epidemiol 189 (2020): 1238–1243.

16. Ashayeri M, Piri S, Abbasabadi N. Evaluating Occupant Feedback on Indoor Air Quality Perception During Covid Stay-at-home Using Social Media Data: A Nationwide Study in the U.S. (2023).

17. Ibrahim AM, Mohammad AH. Assessment of COVID-19 Precautionary Measures in Sports Facilities: A Case Study on a Health Club in Saudi Arabia. J Build Engineering 46 (2022): 103662.

18. Frumkin H. COVID-19, the Built Environment, and Health. Environ Health Perspect 129 (2021): 075001.

19. Roh T, Alejandro M-R, Juha B, et al. Indoor Air Quality and Health Outcomes in Employees Working from Home during the COVID-19 Pandemic: A Pilot Study. Atmosphere 12 (2021): 1665.

20. Kelly FJ, Julia CF. Improving Indoor Air Quality, Health and Performance within Environments Where People Live, Travel, Learn and Work. Atmospheric Environ 200 (2019): 90–109.

21. Nwanaji-Enwerem JC, Joseph GA, Paloma IB. Another Invisible Enemy Indoors: COVID-19, Human Health, the Home, and United States Indoor Air Policy. J Exposure Sci Environ Epidemiol 30 (2020): 773–775.

22. Sheikh KD, Jakub K, Peter W. Application of an Occupant Voting System for Continuous Occupant Feedback on Thermal and Indoor Air Quality –Case Studies in Office Spaces. Energy and Buildings 251 (2022): 111363.

23. Xie R, Yiyang X, Jinhui Y, et al. Indoor Air Quality Investigation of a Badminton Hall in Humid Season through Objective and Subjective Approaches. Sci Total Environ 771 (2021): 145390.

24. Zhou X, Yanfeng L, Cong S, et al. Impact of Temperature and Microstructure on the Emission Characteristics of VOC in Porous Building Materials. Proc Engineering 121 (2015): 1067–1075.

25. Jarnstrom H, Saarela K, Kalliokoski P, et al. Reference Values for Indoor Air Pollutant Concentrations in New, Residential Buildings in Finland. Atmospheric Environ 40 (2006): 7178–791.

26. Tang H, Yong D, Brett C. Post-Occupancy Evaluation of Indoor Environmental Quality in Ten Nonresidential Buildings in Chongqing, China. J Building Engineering 32 (2022): 101649.

27. Guo X, Kyumin L, Zhe W, et al. Occupants' Satisfaction with LEED- and Non-LEED-Certified Apartments Using Social Media Data. Build Environ 206 (2021): 108288.

28. Villeneuve H, Brien WO. Listen to the Guests: Text-Mining Airbnb Reviews to Explore Indoor Environmental Quality. Building Environ169 (2020): 106555.

29. Chinazzo G. Geographical Distribution of the Sources of IEQ Complaints: An Analysis through Text-Mining of Online Job Reviews. J Phys Conf Ser 2069 (2021): 012176.

30. Ashayeri M, Abbasabadi N. Unraveling energy justice in NYC urban buildings through social media sentiment analysis and transformer deep learning. Energy and Buildings 306 (2024): 113914.

31. Silge J, David R. Text Mining with R. https://www.tidytextmining.com/ (2022).

32. Saadoon AH, Usman IMS, Hanafiah MM. et al. Online Review Analysis of Perceived Indoor Environmental Quality (IEQ) at Former Green Certified Hotels in Kuala Lumpur, Malaysia 14 (020).

33. Moezzi M, John G. Text Mining for Occupant Perspectives on the Physical Workplace. Build Res & Information 39 (2011): 169–182.

34. Romero-Silva R, Sander de L. Learning from the Past to Shape the Future: A Comprehensive Text Mining Analysis of OR/MS Reviews. Omega 100 (2021): 102388.

35. David A, Grünwald E, Cichy P, et al. The Application of Text Mining Methods in Innovation Research: Current State, Evolution Patterns, and Development Priorities. R&D Management 50 (2020): 329–351.

36. Debortoli S, Müller O, Iris J, et al. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. Commun Assoc Info Syst 39 (2016): 110–135.

37. Bazzaz A, Kashani MH, Mahdipour E, et al. Big Data Analytics Meets Social Media: A Systematic Review of Techniques, Open Issues, and Future Directions. Telematics and Informatics 57 (2021): 101517.

38. Philipp F. Automation and Autonomy of Big Data-Driven Algorithmic Decision-M Aking, Contemporary Readings in Law and Social Justice 11 (2019): 51–56.

Ashayeri M, et al., J Environ Sci Public Health 2024
DOI:10.26502/jesph.96120205

39. Martí P, Serrano-Estrada L, Nolasco-Cirugeda A. Social Media Data: Challenges, Opportunities and Limitations in Urban Studies. Comput Environ Urban Sys 74 (2021): 161–174.

40. Wang N, Phelan PE, Chioke H, et al. Past Visions, Current Trends, and Future Context: A Review of Building Energy, Carbon, and Sustainability. Renew Sus Energy Rev 82 (2018): 976–993.

41. Kim AE, Heather MH, Murphy J, et al. Methodological Considerations in Analyzing Twitter Data. JNCI Monographs 47 (2013): 140–146.

42. Wickham H, Romain F, Lionel H, et al. Dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org (2022).

43. Wickham H, Maximilian G. Tidyr: Tidy Messy Data. https://github.com/tidyverse/tidyr (2022).

44. Feinerer I, Kurt H. Tm: Text Mining Package. https://CRAN.R-project.org/package=tm (2020).

45. Huang L, Perry PO, Nielsen FA, et al. Corpus: Text Corpus Analysis. https://leslie-huang.github.io/r-corpus/ (2021).

46. Xie Y, Linh Le, Yiyun Z, et al. Deep Learning for Natural Language Processing. In Handbook of Statistics, Elesvier 38 (2018): 317–328.

47. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org (2016).

48. Getting Started with the Twitter API. https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api (2022).

49. Pilař L, Lucie KS, Jana P, et al. Twitter Analysis of Global Communication in the Field of Sustainability. Sustainability 11 (2019): 6958.

50. Barrie C, Chun-ting J. AcademictwitteR: An R Package to Access the Twitter Academic Research Product Track v2 API Endpoint. JOSS 6 (2021): 3272.

51. Proellochs N, Stefan F. SentimentAnalysis: Dictionary-Based Sentiment Analysis (2021).

52. Medhat W, Ahmed H, Hoda K. Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering J 5 (2017): 1093–1113.

53. Tyler WR. Quantitative Discourse Analysis Package. https://github.com/trinker/qdap (2020).

54. Neuwirth ER, ColorBrewer P. Number of U.S. Multifamily Homes 2015-2023 10 (2023).

55. Carroll J, Alicia S, Jonathan S, et al. Easy Access to "ggplot2" Commands (2021).

56. Mousavi A, Jun W. Indoor-Generated PM 2.5 During COVID-19 Shutdowns Across California: Application of the PurpleAir Indoor–Outdoor Low-Cost Sensor Network. Environ Sci Technol 55 (2021): 5648–5656.