Cluster analysis of IAQ patterns and characteristics in a university dormitory

Joongyeon Jo¹, Hyeun Jun Moon¹
¹Dankook University, Yongin, Republic of Korea

Abstract
The importance of indoor air quality (IAQ) research in university dormitories, which serve as living spaces for students, is increasing. In this study, CO₂ and PM2.5 sensors were installed in 10 university dormitory rooms to measure concentration variations. The concentrations of CO₂ and PM2.5 exhibited diverse patterns throughout the day, and due to the nature of dormitory living, significant differences between weekdays and weekends were challenging to identify. Using the K-means++ algorithm, this study categorized the pollution levels of IAQ in university dormitory rooms into three levels: low-, medium-, and high-pollution rooms. It demonstrated that clustering could differ based on pollution sources. Consequently, it was confirmed that both HVAC systems and air purifiers are required for the management of IAQ in university dormitory rooms. The findings could help in managing good indoor environment of dormitories.

Highlights
• Indoor air quality analysis in 10 dormitory rooms in a Korean university using CO₂ and PM2.5 data.
• Clustering methods to categorize private rooms based on CO₂ and PM2.5 concentrations and patterns.
• The findings support previous studies linking indoor air quality to occupancy pattern and provide data for optimizing indoor environments in dormitories.

Introduction
The indoor environment is a crucial component of human health and well-being, given that modern people spend more than 90% of their time indoors. The quality of indoor air, temperature, lighting, and noise affects occupant comfort, productivity, and health. Poor indoor environmental quality (IEQ) has been linked to various health issues, such as allergies, asthma, respiratory infections, and sick building syndrome. In recent years, there has been a growing interest in achieving healthy and comfortable indoor environments, and research on IEQ to achieve this goal is actively being conducted worldwide. In particular, through the COVID-19 pandemic, interest in IAQ is increasing among the elements that make up a healthy indoor environment. To quantify the trends in IAQ research, we analyzed the number of articles related to 'healthy environment' and 'indoor air quality (IAQ) analysis' using the Scopus citation database. Our search included articles published between 2010 and 2022. The results showed a significant increase in the number of articles related to healthy environments, from 1552 in 2010 to 5663 in 2022. Similarly, the number of articles related to IAQ analysis has increased from 66 in 2010 to 276 in 2022. These findings indicate a growing emphasis on IAQ research in achieving healthy and comfortable indoor environments.

Continuous measurement and verification of IAQ is necessary in various domains. This entails (i) enhancing understanding of the determinants of indoor air pollution, such as identifying pollution sources and assessing temporal evolution, (ii) providing valuable information to occupants and building managers regarding the relationship between daily activities and levels of indoor pollution, as well as issuing alerts and implementing corrective actions when exceeding threshold values, and (iii) controlling the operation of ventilation or air purification systems to achieve the best compromise between health and energy consumption considerations (Caron et al). In particular, in the context of university dormitories, where many students reside for extended periods, it is crucial to prioritize special management of IAQ to enhance the students' health and academic performance. However, in most cases, a one-size-fits-all approach in building management that neglects the specific characteristics of each room, such as occupancy patterns and activities, leads to a deterioration in IAQ and energy wastage simultaneously.

In numerous prior studies, in the case of absence of training dataset or prior knowledge about the data's characteristics and structure, the use of unsupervised learning approaches, particularly the centroid-based clustering algorithm, K-means++, has been extensively demonstrated in the field of data mining. This technique has been frequently adopted due to its proven performance and the convenience of implementation and usage, and ability to dealing with datasets that contain missing values (Verma et al.). Given these advantages, we have chosen to utilize this method to investigate the IAQ in university dormitory rooms.

• Background
In Korea, researchers have paid particular attention to vulnerable populations, such as the elderly, infants, and adolescents, who spend more time indoors and have weaker immune systems. Choi (2015) surveyed senior centers and children's centers in Seoul and found that over 9,000 facilities had insufficient consideration for indoor air quality. Although both managers and users were
concerned about indoor air quality, their knowledge on how to manage it was inadequate, resulting in inappropriate responses and practices. Thus, there is a need to establish a system for monitoring indoor air quality management and supporting substandard facilities.

Adolescents have been extensively studied as the next most sensitive group. Lee, Son, and Kim (2010) investigated the concentration of indoor air pollutants and the health effects of volatile organic compounds (VOCs) in schools, while Ryu, Hong, and Kwon (2014) assessed indoor air quality in 60 schools by examining formaldehyde (HCHO) and total volatile organic compounds (TVOCs). They found that new furniture was the primary source of pollutant emissions in the indoor environment. Jeong et al. (2014) analyzed particulate matter (PM10) and total bacteria counter (TBC) concentrations in 110 elementary schools and suggested that the operation of elementary school indoor air quality plans should be preceded by a daily pattern and activity pattern analysis for IAQ assessment. Jeong et al. (2019) conducted an indoor air quality survey in 10 schools and found that 75% of the schools did not meet environmental standards. Notably, the concentration of carbon dioxide (CO2) and PM10 was higher in school facilities than in other public facilities. To improve indoor air quality in schools, Noh et al. (2021) investigated the levels of PM10, PM2.5, and CO2 in classrooms and recommended the development and distribution of customized air purifiers that consider the activities of classroom occupants.

However, previous studies have been conducted on sensitive facilities such as senior centers and daycare centers, and educational facilities such as classrooms, and there is a lack of research on the indoor air quality of university dormitories despite their potential impact on the health and academic performance of students. In this study, we aimed to fill this research gap by conducting an in-depth analysis of the indoor air quality of a university dormitory. Specifically, we installed indoor environmental sensors in the study rooms of all first-floor rooms in a dormitory building at D University in Gyeonggi-do, Korea, and measured CO2 and PM2.5 levels for four months. The results of our analysis and proposed improvement measures serve as a valuable basis for future research aimed at improving the indoor air quality of university dormitories.

### Methodology

- **Target building**

For this study, we selected a dormitory at D University located in Yongin-si, Gyeonggi-do, Korea, to investigate the indoor air quality of a university dormitory which is described in Figure 1. The dormitory has a gross floor area of 13,182 m² on the 10th floor. During the semester (September 1, 2022, to December 14, 2022), each of the 10 dormitories was occupied by four students, and during the end of the semester (December 15, 2022, to December 31, 2022), the number of residents varied from vacant to four. Each room consists of a study room, a bedroom, and a bathroom, with the study room and bedroom measuring 10.24 m² and 10.32 m², respectively. The study room has one 1800 mm × 1400 mm sliding window, the bedroom has a balcony with a window to the outside, and the bathroom has one exhaust fan. Ventilation through drafts is difficult as the front door does not have a door stopper. In both the study and bedroom, the ceilings and walls are covered with paper wallpaper, the floors are made of hardwood flooring, and the furniture such as desks and bookcases are made of wood. Neglect of management for the air conditioning (system air conditioner) and exhaust fan in the restroom was discovered, could adversely affect indoor air quality of the room.

**Figure 1 : Target building**

**Figure 2 : Dormitory floor layout**

### Table 1 : Indoor air quality guidelines

<table>
<thead>
<tr>
<th>Institution</th>
<th>Facility Type</th>
<th>CO2</th>
<th>PM2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROK Ministry of Environment</td>
<td>General facilities</td>
<td>1000 ppm</td>
<td>50 μg/m³</td>
</tr>
<tr>
<td></td>
<td>Vulnerable facilities</td>
<td>1000 ppm</td>
<td>35 μg/m³</td>
</tr>
<tr>
<td>ROK Ministry of Education</td>
<td>General classroom</td>
<td>1000 ppm</td>
<td>35 μg/m³</td>
</tr>
<tr>
<td></td>
<td>Forced ventilation</td>
<td>1500 ppm</td>
<td>35 μg/m³</td>
</tr>
<tr>
<td>USEPA</td>
<td>-</td>
<td>N/A</td>
<td>35 μg/m³</td>
</tr>
<tr>
<td>WHO</td>
<td>-</td>
<td>N/A</td>
<td>35 μg/m³</td>
</tr>
<tr>
<td>ASHRAE</td>
<td>-</td>
<td>1000 ppm</td>
<td>25 μg/m³</td>
</tr>
</tbody>
</table>
Data Collection and Preprocessing
We conducted the data collection of CO\textsubscript{2} and PM2.5 in 10 dormitory rooms on the 3rd floor from September 1, 2022, to December 31, 2022 by 1 minute interval. To measure the actual indoor air quality, we installed an integrated indoor environmental sensor in the study room to measure CO\textsubscript{2} and PM2.5. Prior to utilizing the sensor, the calibration of sensor errors was performed within factory setting for the accuracy and reliability of the collected data. In addition, CO\textsubscript{2} and PM2.5 sensor with accuracy of ±50ppm and ±10μg/m\textsuperscript{3} respectively was employed in this study. The additional information for the installed sensors is shown in Table 2.

Table 2: Sensor specification

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Tech</th>
<th>Measuring range</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO\textsubscript{2}</td>
<td>NDIR</td>
<td>0~5000ppm</td>
<td>±50ppm</td>
</tr>
<tr>
<td>PM2.5</td>
<td>Laser Scattering</td>
<td>0~5,000μg/m\textsuperscript{3}</td>
<td>±10μg/m\textsuperscript{3}</td>
</tr>
</tbody>
</table>

IAQ Standards
Indoor air quality standards vary depending on the regulating bodies, such as the Ministry of Environment of Republic of Korea (ROK) and American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). The recommended IAQ standards for different types of facilities are presented in Table 1. In Korea, the Ministry of Environment oversees most facilities, while the Ministry of Education manages elementary, middle, and higher education facilities. The Ministry of Environment recommends PM2.5 of 50μg/m3 or less for general facilities and 35μg/m3 or less for sensitive facilities, while CO\textsubscript{2} is set at 1000 ppm or less regardless of the facility (ROK Ministry of Environment, 2018). The Ministry of Education recommends 1000 ppm CO\textsubscript{2} and 35μg/m3 PM2.5 for all classrooms and 1500 ppm CO\textsubscript{2} or less if mechanical ventilation is used (ROK Ministry of Education, 2018). The US Environmental Protection Agency (EPA) recommends PM2.5 of 35μg/m3 or less based on a 24-hour average (USEPA, 2023), and the World Health Organization (WHO) recommends PM2.5 of 35μg/m3 or less based on a 24-hour average (WHO, 2014). ASHRAE recommends CO\textsubscript{2} below 1000 ppm and PM2.5 below 25μg/m3 (ASHRAE, 2022). This study evaluated the indoor air quality of university dormitory rooms based on CO\textsubscript{2} and PM2.5 and aimed to meet the standards of all institutions with CO\textsubscript{2} below 1000 ppm and PM2.5 below 25μg/m3. The limitation of the pollutants is visually represented by the red-colored area in Figure 8 and 9. This delineation provides a clear indication of the threshold or boundary beyond which the pollutant levels are considered to exceed acceptable limits.

K-means Clustering
As previously mentioned, this study applied an unsupervised clustering algorithm, K-means, to determine the CO\textsubscript{2} and PM2.5 concentration patterns and characteristics of each dormitory room. The K-means Clustering algorithm is commonly used for grouping input data into K clusters, with the average of the inputs in each cluster serving as the centroid. However, K-means has limitations in forming optimal clusters as the location of initially specified centroids is randomly determined, resulting in different outcomes each time. Additionally, if the distance between centroids is short, classification can be challenging. To overcome these limitations, the K-means++ algorithm was utilized in this study, which was proposed by Arthur (2006) to address the randomization problem of center points. To determine the optimal number of clusters, the Elbow method was employed.

Results and Discussion
In this study, 10 target rooms were selected to measure the IAQ in the dormitory of D University. The measurement results are summarized in Table 3, and the 24-hour concentration changes and average concentrations of indoor air pollutants in all rooms are shown in Figure 8 and 9.

Carbon dioxide (CO\textsubscript{2})
CO\textsubscript{2}, a common indoor pollutant, is generated by human respiration processes. Therefore, CO\textsubscript{2} has been widely utilized in occupancy detection and prediction studies, surpassing other environmental sensors in effective occupancy detection (Candanedo et al., 2017) and prediction (Chen et al., 2018). In this study, CO\textsubscript{2} data profiling was conducted to obtain fundamental insights into the target area and occupants before clustering by analyzing the occupancy patterns of dormitory residents based on days. Figure 4 illustrates this analysis, presenting the average CO\textsubscript{2} concentration patterns at different times from Monday to Sunday using colors such as...
as Salmon, Moccasin, Lightgreen, Skyblue, Thistle, Darkgrey, and Grey. The following observations can be made: 1) Each room's occupancy patterns can be categorized as having similar patterns regardless of weekdays or weekends (Rooms 304, 305, 306, 307, 309, 311, 312), exhibiting some differences in patterns between weekdays and weekends (Rooms 303, 308), and showing variations of the patterns in specific times (e.g., 00:00 to 09:00) by the day of the week (Room 310). However, no distinct pattern differences were found specific to each day of the week in all rooms. Therefore, it can be inferred that due to the nature of university dormitories, the residents generally follow relatively scheduled lifestyles. 2) Room 309 had the lowest occupancy rate during the data collection period, while Room 311 had the highest occupancy rate during the same period. Room 309 recorded a minimum concentration of 578 ppm and a maximum concentration of 997 ppm, whereas Room 311 exhibited a minimum concentration of 982 ppm and a maximum concentration of 1863 ppm, indicating variations in occupancy rates among the rooms. Although CO₂ is not directly toxic, prolonged exposure to high concentrations can cause adverse effects, such as dizziness and decreased work performance, and is associated with various air pollutants. Therefore, it is considered a representative pollutant for indoor air quality. To assess the indoor air quality in terms of CO₂ concentration, we classified the rooms into low-, medium-, and high-pollution rooms based on the average concentration relative to the recommended standard. Specifically, rooms with an average concentration within +10% of the recommended standard were categorized as low-pollution rooms, while those with concentrations exceeding +10% of the recommended standard for some hours but staying within +10% of the recommended standard for most hours were classified as medium-pollution rooms. Rooms with concentrations exceeding +10% of the recommended standard for more than half of the day were classified as high-pollution rooms. This classification can help establish IAQ control standards. To cluster the 24-hour average CO₂ concentration pattern of each room, we used the K-means++ algorithm to improve the result, which solves the center point randomization problem. The optimal K value was selected using the Elbow method, and it was found to be K=3 (Figure 5). The K-means clustering was then performed, and the results are shown in Figure 6.
concentrations up to 2357 ppm (Table 3, Room 307) during some hours of the day, and generally recorded concentrations up to 1400 ppm for 6 hours of the day. Room 311 showed a pattern distinct from the other rooms, with a concentration pattern exceeding 1000 ppm on average and staying below 1000 ppm for less than 4 hours of the day, indicating a significant difference in the level of contamination by ventilation and lifestyle compared to the other rooms. Rooms 309 and 310, while exceeding 1000 ppm for some hours, tended to remain below 1000 ppm for 24 hours on average.

Using the criteria set in this study, rooms were categorized as low, medium, or high pollution rooms. Rooms 303, 304, 306, 308, 309, and 310 were categorized as low pollution (low CO2 emission) rooms, rooms 305, 307, and 312 as medium pollution (medium CO2 emission) rooms, and room 311 as a high pollution (high CO2 emission) room. The clustering analysis was performed with the optimal K=3, and it was found that Room 311, Rooms 309 and 310, and the rest formed separate clusters based on their CO2 concentration levels and patterns (Figure 6). Therefore, it can be concluded that it is possible to cluster rooms based on their CO2 concentration levels and patterns. The detailed results are shown in Figure 8 and Table 3.

- PM2.5

Fine particulate matter with a small diameter requires special attention due to its ability to adsorb harmful gases and heavy metals and penetrate deep into the lungs. Unlike carbon dioxide, which is highly influenced by human occupancy, the concentration of PM2.5 is greatly affected by the behavior of occupants rather than occupancy, necessitating individual measurement, analysis, and management.

To evaluate the indoor air quality of each room with respect to PM2.5, Figure 9 and Table 3 illustrate the change in concentration pattern and average concentration throughout the day, while Figure 7 shows the percentage of the total data that exceeds the standard and clustering analysis for each room. Data show that most rooms maintain an average concentration of 25μg/m3 or less and tend to exceed the threshold in a moment by more than three times (e.g., 97μg/m3, Room 303) at certain times. Although most rooms do not exceed the recommended standard of 25μg/m3 for the average concentration throughout the day, the standard is exceeded for more than one hour and less than three hours during most days. Therefore, a boxplot was used to compare the distribution of PM2.5 data in each room. The plots showed various statistical measures, including the maximum value, upper and lower quartiles, median, and minimum value of the dataset. These comprehensive visualizations enable a in-depth analysis of the data distribution and provide valuable insights into the clustering patterns and overall characteristics of the dataset. As shown in Figure 7, the percentage of data exceeding the recommended standard of 25μg/m3 was displayed at the top of the graph for each room. The optimal K=3 was derived using the Elbow method for the proportion of data exceeding the standard in each room, and the resulting clusters were colored according to the proportion at the top of the graph.

Rooms 310 and 311, which exhibited less than 5% of data exceeding the recommended standard, formed a cluster. Rooms 304, 306, and 312, which had more than 5% but less than 10% of data exceeding the standard, formed another cluster. Lastly, rooms 303, 305, 307, 308, and 309, with more than 10% of data exceeding the standard, were determined to be one cluster. The average PM2.5 concentration in all rooms complied with the threshold of 25μg/m3; however, distinct characteristics were observed among different clusters. 1) The low-pollution rooms exhibited a narrower Inter Quartile Range (IQR) with lower maximum values (mean of 12.5 μg/m3) and median values (mean of 3.7 μg/m3). In contrast, the high-pollution rooms displayed a wider IQR range with higher maximum values (mean of 29.4 μg/m3) and median values (mean of 5.6 μg/m3). This indicates that not only does the high-pollution rooms exceed the IAQ standards more frequently, but even within the acceptable range, the air quality is generally poorer compared to the low-pollution rooms. 2) By referring to Figure 4 and Figure 7 together, it can be observed that Room 311, which maintained the highest occupancy rate for an extended period, belongs to the cleanest PM2.5 cluster. On the other hand, Room 309, which recorded the lowest occupancy rate during the same period, belongs to the most polluted PM2.5 cluster. This suggests that the level of indoor air pollution caused by PM2.5 in the dormitory environment is influenced more by the occupants’ behavior during their stay rather than the duration of their presence.

Conclusion

This study investigated the indoor air quality of private rooms in a university dormitory by installing the indoor air quality sensors in 10 dormitory rooms to collect CO2 and PM2.5 data and analyzed the concentrations of the pollutants. Some individual rooms recorded high concentrations of CO2 and PM2.5 during certain hours, indicating the need for individual management of indoor air quality. Furthermore, the concentration change
patterns throughout the day were analyzed to classify the rooms according to the level of pollution. The results showed that the pattern of pollution sources in each dormitory varies greatly and can be divided into three clusters according to the degree of pollution. Clustering analysis categorized the private rooms into low-, medium-, and high-pollution rooms.

It has been observed that the clustering of CO$_2$ and PM2.5 can be different. Room 311, classified as a high-pollution room of CO$_2$ was categorized as a low-pollution room of PM2.5, while Room 303, 305, 307, 308, and 309, classified as medium-pollution room of CO$_2$, were classified as high-pollution room of PM2.5. Room 309, categorized as a low-pollution room of CO$_2$, was classified as a high-pollution room of PM2.5. Rooms 304, 306, and 312 were classified as medium-pollution room for both CO$_2$ and PM2.5, while only Room 310 was classified as a low-pollution room for both pollutants. From this, it can be concluded that both ventilation systems for CO$_2$ control and air purifiers for PM2.5 control are necessary in university dormitory rooms.

This highlights the importance of managing IAQ in private rooms and suggests individual management is necessary to maintain healthy indoor environments. The findings can be useful in establishing future IAQ control standards for university dormitories and other residential buildings. The study found that pollution levels vary by source, and accurate pollution classification requires measurement of various pollutants.

The findings also support previous studies that behavioral patterns of occupants are closely related to indoor air quality. In conclusion, this study provides significant data that can help optimize the indoor environment of dormitories in the future: This can be utilized for optimal HVAC control for indoor air quality control in dormitories in the future, establishing an energy-saving control strategy while maintaining a comfortable indoor environment. Overall, this study provides valuable information for managing indoor air quality in dormitories and improving the overall indoor environment.

Acknowledgement
This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Ministry of Trade, Industry and Energy(MOTIE) (No. 20212020800120).

References


Table 3: Air pollutant measurement result

<table>
<thead>
<tr>
<th>Room</th>
<th>CO₂ (ppm) Min</th>
<th>CO₂ (ppm) Mean</th>
<th>CO₂ (ppm) Max</th>
<th>PM2.5 (µg/m³) Min</th>
<th>PM2.5 (µg/m³) Mean</th>
<th>PM2.5 (µg/m³) Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>303</td>
<td>420.0</td>
<td>864.3</td>
<td>1998.3</td>
<td>11.98</td>
<td>97.2</td>
<td>9.72</td>
</tr>
<tr>
<td>304</td>
<td>418.1</td>
<td>858.8</td>
<td>1788.6</td>
<td>13.2</td>
<td>79.0</td>
<td>13.1</td>
</tr>
<tr>
<td>305</td>
<td>420.5</td>
<td>934.9</td>
<td>1741.4</td>
<td>1.2</td>
<td>9.14</td>
<td>10.8</td>
</tr>
<tr>
<td>306</td>
<td>431.2</td>
<td>931.7</td>
<td>2245.2</td>
<td>0.00</td>
<td>12.0</td>
<td>9.9</td>
</tr>
<tr>
<td>307</td>
<td>423.1</td>
<td>988.6</td>
<td>2357.1</td>
<td>0.00</td>
<td>93.2</td>
<td>92.1</td>
</tr>
<tr>
<td>308</td>
<td>411.6</td>
<td>723.4</td>
<td>2001.1</td>
<td>1.5</td>
<td>151.8</td>
<td>163.4</td>
</tr>
<tr>
<td>309</td>
<td>405.7</td>
<td>701.8</td>
<td>2125.3</td>
<td>0.00</td>
<td>10.8</td>
<td>8.4</td>
</tr>
<tr>
<td>310</td>
<td>407.8</td>
<td>1412.5</td>
<td>2217.8</td>
<td>0.00</td>
<td>8.4</td>
<td>8.2</td>
</tr>
<tr>
<td>311</td>
<td>423.5</td>
<td>978.6</td>
<td>1634.2</td>
<td>1.0</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>312</td>
<td>409.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 8: CO₂ concentration pattern of each room

Figure 9: PM2.5 concentration pattern of each room