Occupants' Satisfaction with LEED- and Non-LEED-Certified Apartments Using Social Media Data

Xingtong Guo^a, Kyumin Lee^b, Zhe Wang^c, Shichao Liu^a

a. Civil and Environmental Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, 01609, USA

b. Computer Science, Worcester Polytechnic, Institute 100 Institute Road, Worcester, MA, 01609, USA

c. Energy Technology Area, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA, 94720, USA

* Corresponding: Shichao Liu (sliu8@wpi.edu)

Highlights

- 1) Proposed a natural language processing method to assess occupants' satisfaction
- 2) Identified 3 topic clusters (*transportation & location*, *running cost*, and *health & wellbeing*)
- 3) The median satisfaction with LEED and non-LEED apartments is positive.
- 4) A significant but small or negligible uptick in satisfaction with LEED apartments
- 5) There is a weak relationship between rent price and star rating.

Abstract

Leadership in Energy and Environmental Design (LEED) certified buildings aim to offer a sustainable and healthy built environment. Previous studies have shown mixed and inconsistent results on whether occupants in LEED-certified buildings are more satisfying than in non-LEED-certified counterparts. Those studies usually based on surveys or questionnaires for commercial buildings were limited by sample size and pre-defined question structures. Since most people stay longer at home during the COVID-19 pandemic and the trend might continue in the post-pandemic era, assessing the satisfaction with LEED-certified residential buildings benefits future environmental design and certification system development. In this work, we propose a natural language processing-based approach for such assessment. The study collected 16,761 online reviews on 260 LEED-certified apartments and 180 non-LEED-certified-apartments from social media, then applied topic modeling and sentiment analysis to evaluate occupants' satisfaction. Based on topic modeling, we categorized online comments into three topic clusters, 1) *location and transportation*, 2) *running cost*, and 3) *health and wellbeing*. The subsequent sentiment analysis has shown a statistically significant but small or negligible enhancement in the satisfaction occurring in LEED-certified apartments compared to non-LEED-certified ones concerning all the three topic clusters. The “*significant but small or negligible uptick*” has also been found in online star rating and indoor environmental satisfaction. The only exception with a large effect size is lighting that is significantly more satisfying in LEED-certified apartments. Nevertheless, the statistical significance in online star rating disappears when it is normalized by rent price and property house value.

Keywords: *Residential buildings, Post occupancy, Topic modeling, Sentiment analysis, IEQ, COVID*

Graphical Abstract

1. Introduction

Leadership in Energy and Environmental Design (LEED) is one of the most widely used green building rating systems. It was developed by the non-profit U.S. Green Building Council (USGBC). As of 2019, nearly 100,000 projects registered and certified LEED commercial projects [1]. LEED can support all building types, such as offices, schools, hospitals, and homes. It is comprised of 9 credit categories, from regional priority to indoor environmental quality (IEQ). A project pursuing LEED certification can earn one of four LEED rating levels — Platinum (>80 points), Gold (60-79 points), Silver (50-59 points), and Certified (40-49 points)—based on the total points earned across those categories [2].

Occupants' satisfaction with buildings can be attributed to IEQ (e.g., lighting, temperature, air quality), workplace, and building features such as aesthetic appearance, furniture, cleanliness, amount of privacy, and level of personal control [3][4], in addition to running energy cost [5]. The largest database of occupant indoor environmental quality survey by the Center for the Built Environment (CBE) focuses on seven areas of indoor environmental performance. It has been implemented in more than 1000 buildings with over 100,000 individual occupant responses as of March 2017 [6]. Based on the analysis of a subset of the dataset, office buildings with LEED certification outperformed non-LEED-certified buildings in occupants' satisfaction regarding building overall, cleanliness, colours and textures, and air quality, even though the effect sizes of the difference was negligible [4]. Using the same dataset, Lee and Kim [7] concluded that LEED-certified buildings received higher satisfaction with office furnishings, thermal comfort, indoor air quality, and cleanliness and maintenance but lower with office layout, lighting, and acoustics. The CBE database has revealed marginal advantages of LEED certification in promoting occupants' satisfaction in office buildings.

Besides the results from the CBE database, Table 1 summarizes the previous nine studies on occupants' satisfaction with a variety of factors associated with LEED-certified and non-LEED-certified buildings. Overall, mixed reported findings have been observed on whether LEED-certified buildings are more satisfying or not, which could be attributed to the differences in building locations, building ages, occupancy period, samples size, or building green features [8]. For instance, teachers in LEED-certified school buildings had a higher satisfaction rate in lighting, thermal comfort, indoor air quality but were less satisfied with acoustics than those in school buildings without LEED certification [9]. However, LEED-certified hospitals have produced elevated satisfaction in terms of all IEQ factors based on a quite small sample size [10].

Four of the nine studies in Table 1 focused on office buildings, but only one study has been reported on residential apartments in terms of occupants' satisfaction with indoor air quality [11]. People spend nearly 90% of their time indoors [12] with half of those time spent at home [13][14]. The role of residential buildings has become even more crucial during the COVID-19 pandemic since many people are forced to work home remotely for a prolonged period due to a virus lockdown. Occupants' satisfaction with residential buildings has become more important than ever before.

All the reviewed studies in Table 1 relied on questionnaires to obtain occupants' satisfactory level. The design of those questionnaires typically adopts a top-down approach that leans toward the perspectives of designers, researchers, and policymakers as opposed to occupants. In particular, the structured questions usually have challenges/limitations to reveal occupants' attitudes on aspects not included in the questionnaires, not to mention response rate, timeliness, and longitudinal tracking. Second, it is cost-prohibitive to survey a substantial number of buildings,

This paper is under review

especially residential ones, using questionnaires. Unlike surveying occupants in large commercial buildings where building managers can distribute questionnaires altogether to hundreds of occupants, reaching out to the occupants in residential buildings (e.g., multi-family apartments) is considerably difficult.

One way to avoid those challenges is the bottom-up method. There are tons of information on the Internet contributed by occupants regarding the satisfaction with buildings in the format of star ratings and written comments (Figure 1). Although the comments are unstructured, the information can be processed using text-mining techniques such as Latent Dirichlet Allocation (LDA) [15], and can shed light on occupants' satisfaction with residential buildings.

Text mining and sentiment analysis have been widely applied in many areas to analyze users' satisfaction and attitudes. Berezina et al. [16] used a text-mining method to understand what factors may satisfy or dissatisfy hotels customers; Villeneuve et al. [17] employed a text-mining method to study the sentiment feelings of IEQ issues of Airbnb guests; Kar [18] investigated factors affecting user satisfaction in mobile payments based on Twitter tweets through sentiment analysis and topic modeling using LDA. The primary advantages of using social media data are two-folds, 1) a substantial amount of public data with significantly more reviews compared to distributed questionnaires, and 2) occupants' open-ended comments with diminished biases compared with predefined structured questions in the questionnaires.

The objective of this study is to compare occupants' satisfaction between LEED-certified and non-LEED-certified apartments in the United States through topic modeling and sentiment analysis of publicly posted comments on social media. We selected apartments as the target building type because more online comments are available for apartments than other types of residential buildings such as single-family houses. Our goal is to shed light on the following research questions:

1. Do LEED-certified apartments have a higher star rating than non-LEED-certified apartments?
2. Which latent topics are the most popular and of interest to occupants based on their online comments?
3. How does occupants' satisfaction vary for different factors (e.g., IEQ, running cost) of the apartments both with and without LEED certification?
4. Would rent price and land value affect occupants' satisfaction in LEED-certified and non-LEED-certified apartments?

The subsequent sections include Methodology to retrieve apartment characteristics from USGBC and online review from social media, and to conduct statistical analysis, Results and Discussions to present the major findings to address the above-mentioned research questions, and Conclusions.

Table 1. Prior studies on occupants' satisfaction with LEED-certified and non-LEED-certified buildings.

Studies	Sample size	Data collection method	Building function and Location	Findings	Significance level of differences ($p < 0.05$)
[19]	15 LEED-certified buildings, 6 self-nominated green buildings, and 160 non-LEED-certified buildings	Online questionnaire (Respondents: $N = 33285$ totally)	Offices, USA	LEED-certified office buildings performed better in most aspects of IEQ, but there was no significant difference in lighting and acoustic quality between LEED-certified and non-LEED-certified buildings.	Statistically significant differences in satisfaction with most aspects of IEQ.
[9]	3 LEED-certified schools, 10 conventional schools, and 20 energy-retrofitted schools	Questionnaire (Respondents: $N = 103$ totally)	Schools, Canada	LEED-certified school buildings performed better in most aspects of IEQ but worse in acoustics.	Statistically significant differences in teachers' satisfaction with IEQ.
[20]	12 LEED-certified offices and 12 conventional offices	Online questionnaire (Respondents: $N = 2545$ totally for core the survey module)	Offices, Canada and the north USA	LEED-certified office buildings performed better in most aspects of IEQ and well-being, but worse in acoustics and lighting quality.	Statistically significant differences in overall environmental satisfaction, satisfaction with noise from HVAC, thermal preference, and visual and physical symptom frequency.
[7]	15 LEED-certified buildings and 200 conventional buildings	Online questionnaire (Respondents: $N = 3769$ for LEED, $N = 36719$ for non-LEED)	Offices, USA	LEED-certified buildings performed better in most aspects of IEQ but worse in lighting and acoustics quality.	N/A

This paper is under review

[10]	Two LEED-certified hospitals and one conventional hospital	Questionnaire (Respondents: $N=54$ for LEED, $N=25$ for non-LEED)	Health care facilities, USA	LEED-certified hospitals performed better in all aspects of IEQ than the conventional hospital without LEED certification.	Statistically significant differences in the satisfaction with IEQ.
[11]	18 LEED-certified apartments and 13 conventional apartments; 61 home visits	Questionnaire (Respondents: $N=37$ totally)	Apartments, USA	LEED-certified buildings performed better in most aspects of indoor air quality.	A statistically significant difference in the perception of stuffy air, observation of pests and inadequate ventilation.
[4]	65 LEED-certified office buildings and 79 non-LEED-certified office buildings	Online questionnaire (Respondents: $N=10129$ for LEED, $N=11348$ for non-LEED)	Mainly for offices, USA	LEED-certified buildings performed better in air quality, building maintenance, colours and textures, and cleanliness but worse in amount of light, ease of interaction, visual privacy, visual comfort, amount of space, noise, and sound privacy. However, the effect sizes are negligible.	Statistically significant differences have been reported in all investigated factors except for temperature, furniture adjustability, and comfort of furnishing.
[21]	One LEED-certified mix-used building, one conventional mix-used building	Questionnaire (Respondents: $N=53$ for LEED, $N=72$ for non-LEED)	Mix-used building, China	LEED-certified buildings performed better in summer temperatures and overall IEQ satisfaction but worse in lighting, noise, and temperatures in winter.	Statistically significant differences in the satisfaction with temperature and lighting. No significant difference in the noise satisfaction.
[22]	One LEED-certified factory and one non-LEED-certified factory	Questionnaire (Respondents: $N=35$ for LEED, $N=35$ for non-LEED)	Factories, Sri Lanka	LEED-certified factory performed better in views to outside, cleanliness, furniture, privacy, and lighting, while worse in thermal comfort, provision of ventilation	Statistically significant difference in thermal comfort, provision of ventilation of work, having control over indoor

for work, and having control over indoor environment.

environment, views to outside, cleanliness, furniture, privacy, and lighting.

Figure 1. Online reviews (star rating and written comments) on apartments from social media; (a) Yelp.com; (b) Apartmentratings.com.

2. Methodology

Figure 2 depicts the workflow of the methodologies in this work including 1) acquisition of online ratings and comments from social media, 2) data cleaning with regular expression (Regex) matching [23] and stop-words removal [24], 3) topic modeling to extract keywords and latent topics to classify sentences, 4) sentiment analysis at the sentence level, and 5) hypothesis testing on the difference in sentiment values (indicators of satisfaction) between LEED-certified and non-LEED-certified apartments.

Figure 2. Methodology workflow

2.1 Data acquisition

We identified commercial apartments in the United States by applying a filter of project type “Multi-Family Residential: Apartment” in the database of USGBC [25]. There were a total of 490 LEED-certified and 794 non-LEED-certified residential apartments in the United States until October 2018. The non-LEED-certified apartments have applied for certification but failed. For each building, the database provides address, project name, LEED system version, rating level (if certified), certification date (if certified) and other information. The database can be publicly accessed on an online repository (<http://dx.doi.org/10.17632/hw59ryytdf.1>). Next, we searched each building on three popular apartment review websites (*Yelp.com*, *Apartmentratings.com*, and *Apartment.com*) based on the address and project name, and then scraped occupants’ posted comments and rating score (from 1 = worst to 5 = best) using a developed web crawling tool or manually (if necessary). The search resulted in 8,230 online reviews (1,182,531 words) for 260 LEED apartments and 8,531 online reviews (1,284,763 words) for 180 non-LEED apartments matching the building projects in the LEED database (Table 2). Each review data point included star rating, descriptive written comments, and date of the review.

This paper is under review

Table 2. Statistics of the comments

	# of apartment	# of comments	Average # of sentences per comment	Average # of words per comment	Average # of words per sentence
LEED	260	8230	10	144	15
Non-LEED	180	8531	11	151	14

2.2 Data pre-processing

Online comments sometimes contain less important or noisy information, so data pre-processing step is required before conducting an in-depth analysis. Our data pre-processing step removes the following information:

Information such as symbols and URL links that were removed by Regular Expression (Regex) matching.

Stop words, like “a”, “is”, “the”, “do” etc. that were taken out by employing the Natural Language Toolkit (NLTK) stop words[26] list. Stop words do not express any emotion [27] or satisfaction but can affect topic modeling results and an optimal topic number [28].

2.3 Topic modeling

Topic modeling is a widely used tool to mine data, analyze latent semantic structures, and find topic in unclassified texts [15]. In this study, we employed LDA method, a generative probabilistic model of corpus [15]. LDA is the most widely used topic modeling approach [29], which adopts an unsupervised learning technique to uncover hidden topic information of a collection of messages. LDA method applies a hierarchical structure of “document-topic-word.” Each document (w) is considered as a probability distribution over topics; and each topic is a probability distribution over words.

Figure 3 shows the graphical model representation of the LDA model. When performing topic modeling, the topic number (k) is an unknown value to be defined by the user. The selection of a proper k is a steppingstone for topic modeling since different k values may lead to different topic results.

One metric to optimize k is to minimize perplexity that measures the quality of latent topics. A lower perplexity value implies a higher model conditional likelihood, in other words, better model [30]. For instance, Chen et al. [31] and Ghosh [32] found an ideal k in the range of 30 to 50 with the lowest perplexity. However, the selection of k cannot be solely relied on perplexity since it only crudely indicates an acceptable amount of topic loss. In addition, topic interpretation from a specific discipline (e.g., social science) should be incorporated in the selection process[33]. In this study, we chose k according to both perplexity and qualitative exploration on comprehensive information coverage based on our best understanding of LEED and non-LEED buildings.

Figure 3. Graphical model representation of topic modeling of online reviews using LDA. M denotes the number of documents; α is the parameter of the Dirichlet prior on the per-document topic distribution; β is the parameter of the Dirichlet prior on the per-topic word distribution; θ is the topic distribution for a document; ϕ is the word distribution for the topic; Z is the topic for a word in the document; $W_{n,v}$ is the specific word. The circles in blue shades represent the distribution of each latent topic.

Preliminary modeling results contained words like “favourite”, “name”, “anyone” which may be irrelevant to occupants’ satisfaction. These suspiciously irrelevant words were removed when the coherence score, which can identify if a topic is semantically interpretable [34], resulted from the topic modeling increased without those words.

2.4 Online comments classification using supervised learning

Since a sentence of online comments may contain mixed topics and sentiments, they need to be classified into each founded latent topic before sentiment analysis. As for sentence classification, we used a seed word dictionary as described below, and matched the dictionary with online comments. If a sentence referred several topics, it will be classified into all relevant topics rather than be counted only once. There are several ways to build a dictionary. The popular method, a heuristic approach of combining human annotation and Wordnet’s Synset, sets of cognitive synonyms expressing a particular concept [35]. But it doesn’t work in our case due to professional specificity. For example, when we try to find Synset of “air”, the Wordnet will find words like “bare”, “beam”, “line” and many other words irrelevant to an apartment. Therefore, we build a relevant expanded word dictionary by selecting topic relative words from online comments and LEED reference guides [36][37][38][39]. The detailed dictionary is attached in the Appendix.

2.5 Sentiment analysis

Sentiment analysis, also known as opinion mining, is a process of understanding written contexts and is generally used to determine whether the context contains positive, neutral or negative opinions [40][41][42]. Sentiment analysis in this work is a probabilistic supervised machine learning approach based on the Naïve Bayes algorithm as follows.

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} \overbrace{P(d|c)}^{\text{likelihood}} \overbrace{P(c)}^{\text{prior}} \quad (1)$$

$$C_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{f \in F} P(f|c) \quad (2)$$

Here C denotes a set of all possible classes (negative, neutral, positive) and c is one of the classes; d denotes a document (each sentence of a comment from social media). F means all features value pairs (e.g., location-near, cost-high) and f is one of these feature pair.

The Naïve Bayes classifier performs well when the output is categorical, while it may do poorly for regression problems by discretizing the target value [43]. To overcome the limitation and guarantee the accuracy of the Naïve Bayes model, we used the following Bayes equation for regression:

$$p(Y|E) = \frac{p(E,Y)}{\int p(E,Y)dY} = \frac{p(E|Y)p(Y)}{\int p(E|Y)p(Y)dY} \quad (3)$$

where the likelihood $p(E|Y)$ is the probability density function (pdf) of the evidence E for a given target value Y , and the prior $p(Y)$ is the pdf of the target value before any evidence has been seen.

In this study, the sentiment analysis was conducted with NLTK and the sentiment analysis application programming interface (API) provides polarity values. The final output was a floating-point value from -1 (negative sentiment) to +1 (positive sentiment). In this study, we considered binary classifications, namely, a negative sentiment as dissatisfaction and a positive value to be satisfaction. A higher absolute sentiment value implies that the occupants are more satisfied or dissatisfied with their apartment.

2.6 Statistical analysis

The initial data analysis with the Shapiro-Wilk normality test showed that all datasets were non-normally distributed. Therefore, we assessed the difference of the median with the Wilcoxon signed-rank test. The statistical significance was based on $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). Cohen's d was used to calculate the effect size of the difference that $|d| < 0.147$ “negligible”, $|d| < 0.33$ “small”, $|d| < 0.474$ “medium”, otherwise “large”, following our previous method [44].

3 Results and discussions

In this section, we display a geographical map on the distribution of LEED and non-LEED multi-family residential apartments reviewed in social media, in addition to occupants' satisfaction based on online star rating and sentiment analysis. This section also reports occupants' topics of interest. Besides, the satisfaction with indoor air quality, thermal comfort, acoustics, lighting, and layout is discussed by comparing with prior studies.

3.1 Geographical distribution of LEED and non-LEED multi-family residential apartments

Figure 4 shows the distribution of 260 LEED-certified and 180 non-LEED-certified apartments assessed in this study. The background colors show the median value of housing price in 2018 [45]. Both LEED and non-LEED certified apartments that applied for but did not obtain certification are more concentrated in coastal states such as California and New York. Those states are generally populous and ambitious to achieve sustainability and environmental goals by motivating developers to voluntarily pursue third party certifications for their real estate projects [46].

Figure 4. Distribution of LEED-certified and non-LEED-certified multi-family residential apartments reviewed on social media. The numbers in each state represent # of non-LEED apartments (# of LEED apartments).

3.2 Occupants' satisfaction with LEED-certified and non-LEED-certified residential apartments

The rating of online reviews in most social media ranges from one (worst) to five (best) stars. In this work, we coded the star rating using a 5-scale Likert scale, very dissatisfied (1 star), dissatisfied (2 stars), neutral (3 stars), satisfied (4 stars) and very satisfied (5 stars) to facilitate the analysis of occupants' satisfaction. Figure 5 shows the percentage of the different satisfaction levels for the LEED-certified and non-LEED-certified apartments.

Figure 5. Star ratings of LEED-certified and non-LEED-certified apartments

The total percentage of reviews with satisfied (4 stars) and very satisfied (5 stars) is 69.3% for LEED-certified apartments and 56.4% for non-LEED-certified apartments. The median ratings are very satisfied (5 stars) for LEED-certified and satisfied (4 stars) for non-LEED-certified apartments. The Wilcoxon signed-rank test reports a statistically significant difference ($p < 0.001$) between the median rating of the two apartment types, though the effect size of the difference is negligible according to Cohen's d ($|d| < 0.33$). In other words, LEED-certified multi-family residential apartments are perceived slightly more satisfying than non-LEED-certified apartments according to online star ratings. However, the crude rating results could not reveal how occupants feel about specific aspects of the apartments, which necessities detailed analysis of the contexts and topics of posted comments.

3.3 Topic modeling

The latent semantic structures of occupants' online comments can be revealed through topic modeling. A "topic" in topic modeling is a cluster of words that co-occur in the same documents according to certain patterns through unsupervised learning. We identified the number of topics or themes discussed in those comments based on the perplexity calculated with Gensim [47]. Perplexity measure is a commonly used computational metric to determine the number of topics summarizing an online review [15]. The optimal number of topics resulted from LDA model exists when the perplexity value is the minimum [15]. We calculated the perplexity value for the topic number varying from 1 to 10 and found that the local minimum perplexity occurred when the topic number was equal to 3 (Figure 6).

Despite the advantages of using perplexity to determine the correct topic number, however, it is worth noting that the calculation of perplexity only creates an initial selection of models with an acceptable amount of information loss [33]. Rather, the selection should rely on the best interpretability of the topics. When the topic number is two to four, the topic modeling by the LDA shows that important keywords are missing regarding occupants' satisfaction with apartments even though the perplexity value is small. It is likely that computational algorithms (e.g., LDA) identify nuances that are not semantically meaningful [48]. Therefore, considering both perplexity and

interpretability, we found that five topics can generate comprehensive and semantically meaningful information through topic modeling.

Figure 6. Perplexity of topic modeling for various topic numbers

Table 3 lists ten keywords and their weights that were calculated based on frequency and topic relevancy extracted from LDA for each of the five topics. According to the keywords, totally six themes can be summarized, 1) Location and transportation, 2) IEQ, 3) Neighbourhood, 4) Management service, 5) Running cost, and 6) Amenities. A topic can be a cluster of multiple themes due to the unstructured characteristics of occupants' open-ended comments. For instance, Topic 1 is composed of location and transportation ("car"), pet-policy ("dog"), IEQ ("floor", "hear", "door", "wall"), neighbour ("people", "neighbour"). On the other hand, a theme may appear in different topics. Apparently, these produced themes by topic modeling do not cover the entire spectrum of credit categories specified by either LEED or other rating systems, such as Building Research Establishment Environmental Assessment Method (BREEAM) in Table 4.

Table 3. Keywords of topics extracted from comments

Themes of Topic 1 (Location and transportation, Pet-policy, IEQ, Neighbourhood)		Topic Themes 2 (Management service, Running cost, Neighbourhood)		Topic Themes 3 (Running cost, Management service, Neighbourhood)		Topic Themes 4 (Location and transportation, Running cost)		Topic Themes 5 (Management service, Location and transportation, Amenity, IEQ)	
Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight
people	0.019	apartment	0.039	rent	0.033	parking	0.094	staff	0.037
dog	0.018	maintenance	0.026	pay	0.027	free	0.028	friendly	0.030
apartment	0.018	staff	0.024	management	0.023	park	0.024	location	0.030
floor	0.013	move	0.023	tenant	0.017	store	0.021	apartment	0.030
hear	0.012	resident	0.023	water	0.014	spot	0.023	area	0.030
door	0.012	live	0.023	lease	0.013	food	0.015	building	0.026
building	0.012	property	0.018	money	0.011	shop	0.015	live	0.025
car	0.009	management	0.016	trash	0.009	shopping	0.012	amenity	0.019
neighbour	0.009	service	0.015	unit	0.009	restaurant	0.012	clean	0.018
wall	0.009	office	0.014	office	0.009	close	0.011	build	0.017

LEED defines 9 credit categories for building rating, location and transportation, IEQ, energy, and atmosphere [49], just to name a few. Similarly, BREEAM's criteria [50] consist of 10 categories such as management, health and wellbeing, energy and transportation, as described thoroughly in Table 4. The misalignment between the clustered topics generated by topic modeling and credit categories of the established rating systems (e.g., LEED, BREEAM) might be caused by the fact that occupants are generally unaware of factors other than what they can perceive directly. In other words, despite the potential benefits of sustainability, efficiency and health advocated by those rating systems, occupants may not appreciate the claimed value-adding in general because their issues of concern might not align completely with the scope of the abovementioned rating systems. Many categories such as regional priority and sustainable sites have not been mentioned or implied in occupants' online comments. Nevertheless, the modelled topics by topic modelling and credit categories of LEED and BREEAM still overlap on three clusters of topics: 1) *location and transportation*; 2) *running cost* (mainly refers to rent, maintenance, bills); and 3) *health and wellbeing*. For instance, the category of *IEQ* specified in LEED and summarized by topic modeling, and *pollution* in BREEAM can be clustered into *health and wellbeing*. The measure to assemble topics in three major clusters can facilitate the comparison of occupants' satisfaction with between LEED-certified and non-LEED-certified apartments.

Table 4. Credit categories for LEED and BREEAM

LEED [49]		BREEAM[50]	
Categories	Credits	Categories	Credits
Integrative Process	1	Management	19
Location and Transportation	15	Transportation	12
Sustainable Sites	10	Land use and Ecology	up to 13
Water Efficiency	12	Water	9
Energy and Atmosphere	34	Energy	31
Materials and Resources	13	Materials	12
IEQ	16	Health and Wellbeing	up to 20
Innovation	6	Innovation	up to 2
Regional Priority	4	Pollution	up to 17
		Waste	11

Figure 7 summarizes the distribution of occupants' topics of interest characterized by percentage of sentences discussing a topic cluster for both LEED-certified and non-LEED-certified apartments. Occupants in LEED-certified apartments appear to be more (61.0% vs 56.3%) attentive to factors pertaining to "*Health and wellbeing*" than those in non-LEED-certified counterparts. A possible reason is that the topic cluster covers more themes such as "*amenities*," "*management services*," "*pet-policy*," "*appliance*," and "*indoor environment*" (Table 2) and is related to a larger sample of words like "*noise*," "*air*," "*view*" and "*clean*" than the other two topic clusters. Nevertheless, it is surprising that "*running cost*" has been discussed the least of time, only 11.7% for LEED-certified and 12.5% for non-LEED-certified apartments.

Figure 7. Occupants' topics of interest based on the percentage of sentences discussed in social media (totally 82,890 sentences for LEED apartments and 92,854 sentences for non-LEED-certified apartments)

The top three weighted keywords in the online review for LEED-certified apartments are “staff”, “friendly”, “area” as opposed to “staff”, “management”, and “parking” for non-LEED-certified apartments (Figure 8). The heat maps indicate that online reviews primarily focus on apartment management service, which is supported by the commonly appeared words (e.g., “staff” and “location”) based on frequency for both apartment types. Nevertheless, leisure facilities are discussed more often for LEED-certified apartments. For example, “amenities”, “pool”, and “gym” appear more frequently in those apartments. However, “maintenance” is a popular topic in online reviews for only non-LEED-certified apartments. Furthermore, the heat maps show no clear trend of weight change for most keywords.

Occupants' satisfaction with the three topic clusters regarding LEED-certified and non-LEED-certified apartments can be revealed through sentiment analysis of their online review comments. Figure 9 shows the distribution of the sentiment values of all sentences for each topic cluster between LEED-certified and non-LEED-certified apartments. Generally, occupants have been satisfied with the three topic clusters for both apartment types, as indicated by the slightly positive median sentiment values ranging from 0 to 0.25. Occupants are more satisfied with "*location and transportation*," "*running cost*" and "*health and wellbeing*" for LEED-certified apartments than those without LEED certification. While the differences are statistically significant ($p < 0.001$) in terms of all the three topic clusters, the effect sizes of those differences are negligible according to Cohen's d ($|d| < 0.147$). It is therefore concluded that LEED-certified apartments generate negligibly higher satisfaction with "*location and transportation*," "*running cost*," or "*health and wellbeing*."

Figure 9. Sentiment analysis of online comments (statistical significance: *** $p < 0.001$)

3.5 Satisfaction with various IEQ factors

IEQ is an essential element contributing to people's health and wellbeing. Occupants' sentiment with IEQ factors, such as thermal comfort, acoustics, lighting, indoor air quality, layout, inevitably influences their general satisfaction with the topic cluster "health and wellbeing." Figure 10 shows that LEED-certified buildings outperform non-LEED-certified counterparts for all the five IEQ factors generated by topic modeling, with different significant levels and effect sizes. The large distinction in satisfaction occurs for lighting only. It is observed that acoustics is the least satisfying while thermal comfort receives the highest satisfaction rate in apartments both with and without the LEED certification. Similar findings were also reported previously [13].

Furthermore, we compared the results from sentiment analysis with four previous studies that were conducted in buildings with different functions such as houses and office (Figure 10). The original scale (-1 to 1) of the sentiment values was adjusted linearly to -3 (very dissatisfied) to 3 (very satisfied) to facilitate the comparison. Albeit the differences in sample size and methodology. The results in Figure 10 suggest that unstructured online comments from social media can be applied to analyze occupants' satisfaction through semantic analysis.

Figure 10. The comparison of occupants' satisfaction with IEQ in various building types using different methods. The statistical tests were conducted based on the data of this study only.

Statistical significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, the * in red refers to a large effect size based on Cohen's d ($|d| < 0.147$ "negligible", $|d| < 0.33$ "small", $|d| < 0.474$ "medium", otherwise "large")

3.6 Rent price, land value, and star rating

LEED-certified apartments are expected to be more environmentally friendly and energy efficient [51][52], due to strict construction standards and high quality building materials (low-emitting, recyclable, etc.) [49]. LEED certification awards buildings in the location promoting less vehicle travel distance and better liveability [53]. Accordingly, those locations are generally characterized with a high land property value (Figure 11a). The difference between the land property value of the two apartment types is statistically significant and with a medium effect size (Cohen's d). Surprisingly, the increment in the property value does not necessarily result in an elevation in the rent price (normalized by the bedroom number) for LEED-certified apartments. Figure 11b depicts that the median rent prices per room are not statistically different between apartments with and without LEED certification.

Figure 11. (a) Median house property values and (b) Rent price per room for LEED-certified and non-LEED-certified apartments. The median house property value (in 2018) was retrieved from Federal Housing Finance Agency (FHFA) [54] based on zip code.

Impact of rent price and land property value on star rating

To study how much satisfaction people can get for every dollar they spent, we normalized the star rating (from 1 to 5 stars) of each apartment by the rent price per room to monetize occupants' overall satisfaction. In Figure 12, the normalized star rating of LEED-certified apartments is not statistically ($p = 0.073$) higher than that of non-LEED-certified apartments, indicating that occupants in the LEED-certified apartment are not statistically more satisfied as for the price they pay for the rent. Besides, the correlations between star rating and normalized price of LEED-certified and non-LEED-certified apartments are -0.04 and 0.16 separately, suggesting a weak relationship between rent price and star rating.

Figure 12. Normalized star ratings by rent price per room for LEED-certified and non-LEED-certified apartments

3.7 Limitation

A major limitation of this study could be sample bias. The occupants' online review in this study were collected from *Yelp.com*, *Apartmentratings.com*, and *Apartment.com* before October 2018 when the data analysis was started. Further online review data since then were not considered in this work. Even though the sample size is way larger than that of conventional questionnaires and the reviews from the three websites accounted for most publicly available information on social media, the dataset may not represent the satisfaction of a larger or a different population. For instance, seniors with limited resources may have difficulties posting their comments online. And there are possibilities that some comments are not written by occupants. In addition, the database in this study only includes rental apartments.

4. Conclusions

The goal of this study is to compare occupants' satisfaction level of LEED-certified and non-LEED-certified apartments by analyzing online reviews posted on social media with natural language processing (NLP) since the conventional post occupancy evaluation (POE) for this purpose is generally limited by sample size and pre-defined question structure in questionnaires.

The online review data regarding apartments can be categorized into three topic clusters, 1) *location and transportation*; 2) *running cost* and 3) *health and wellbeing*. Occupants have discussed *health and wellbeing* more frequently (accounting for 56-61%) on social media. Facilities for leisure (e.g., pool, gym) are discussed more often for LEED-certified apartments compared to non-LEED-certified ones.

The sentiment analysis shows that both apartment types have positive median sentiment values for all the three topic clusters, indicating that occupants are satisfied with their apartments in general.

Overall, LEED-certified apartments have significantly higher satisfaction than non-LEED-certified counterparts for most investigated perspectives. However, the enhancement is mostly

This paper is under review

negligible or small according to the calculated effect sizes. In particular, the significant but negligible or small uptick has been found in 1) online holistic star rating, 2) sentiment values of all the three topic clusters, and 3) satisfaction with IEQ factors except for lighting. When the star rating is normalized by the rent price and house property value, no statistical difference ($p = 0.073$) can be found between the two apartment types.

By comparing the evaluation of occupants' satisfaction using online reviews on social media to previous studies with questionnaires, we found that topic modeling and sentiment analysis can be applied to evaluate occupants' satisfaction with buildings.

Acknowledgment

This work was supported by U.S. National Science Foundation (NSF) under Grant No. 2028224.

Declaration of Competing Interest

The authors declare no competing interests.

Reference

- [1] No title, (n.d.). <https://www.usgbc.org/articles/leed-reaches-new-milestone-surpasses-100000-commercial-green-building-projects>.
- [2] No title, (n.d.). <https://www.usgbc.org/leed>.
- [3] M. Frontczak, S. Schiavon, J. Goins, E. Arens, H. Zhang, P. Wargocki, Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design, *Indoor Air*. 22 (2012) 119–131. <https://doi.org/10.1111/j.1600-0668.2011.00745.x>.
- [4] S. Altomonte, S. Schiavon, Occupant satisfaction in LEED and non-LEED certified buildings, *Build. Environ.* 68 (2013) 66–76. <https://doi.org/10.1016/j.buildenv.2013.06.008>.
- [5] K. Amasyali, N.M. El-Gohary, Energy-related values and satisfaction levels of residential and office building occupants, *Build. Environ.* 95 (2016) 251–263. <https://doi.org/10.1016/j.buildenv.2015.08.005>.
- [6] No title, (n.d.). <https://cbe.berkeley.edu/research/occupant-survey-and-building-benchmarking/>.
- [7] Y.S. Lee, S. Kim, Indoor environmental quality in LEED-certified buildings in the U.S., *J. Asian Archit. Build. Eng.* 7 (2008) 293–300. <https://doi.org/10.3130/jaabe.7.293>.
- [8] M. Khoshbakht, Z. Gou, Y. Lu, X. Xie, J. Zhang, Are green buildings more satisfactory? A review of global evidence, *Habitat Int.* 74 (2018) 57–65. <https://doi.org/10.1016/j.habitatint.2018.02.005>.
- [9] M.H. Issa, J.H. Rankin, M. Attalla, A.J. Christian, Absenteeism, performance and occupant satisfaction with the indoor environment of green Toronto schools, *Indoor Built Environ.* 20 (2011) 511–523. <https://doi.org/10.1177/1420326X11409114>.
- [10] S.K. Kim, Y. Hwang, Y.S. Lee, W. Corser, Occupant comfort and satisfaction in green healthcare environments: A survey study focusing on healthcare staff, *J. Sustain. Dev.* 8 (2015) 156–173. <https://doi.org/10.5539/jsd.v8n1p156>.

- [11] M.D. Colton, P. Macnaughton, J. Vallarino, J. Kane, M. Bennett-Fripp, J.D. Spengler, G. Adamkiewicz, Indoor air quality in green vs conventional multifamily low-income housing, *Environ. Sci. Technol.* 48 (2014) 7833–7841. <https://doi.org/10.1021/es501489u>.
- [12] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J. V Behar, S.C. Hern, W.H. Engelmann, The national human activity pattern survey, *Lawrence Berkeley Natl. Lab.* 11 (2011) 231–252. <http://exposurescience.org/the-national-human-activity-pattern-survey-nhaps-a-resource-for-assessing-exposure-to-environmental-pollutants>.
- [13] M. Frontczak, R.V. Andersen, P. Wargocki, Questionnaire survey on factors influencing comfort with indoor environmental quality in Danish housing, *Build. Environ.* 50 (2012) 56–64. <https://doi.org/10.1016/j.buildenv.2011.10.012>.
- [14] E. Lee, Indoor environmental quality (IEQ) of LEED-certified home: Importance-performance analysis (IPA), *Build. Environ.* 149 (2019) 571–581. <https://doi.org/10.1016/j.buildenv.2018.12.038>.
- [15] D.M. Blei, A.Y. Ng, M.T. Jordan, Latent dirichlet allocation, *Adv. Neural Inf. Process. Syst.* 3 (2002) 993–1022.
- [16] K. Berezina, A. Bilgihan, C. Cobanoglu, F. Okumus, Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews, *J. Hosp. Mark. Manag.* 25 (2016) 1–24. <https://doi.org/10.1080/19368623.2015.983631>.
- [17] H. Villeneuve, W. O'Brien, Listen to the guests: Text-mining Airbnb reviews to explore indoor environmental quality, *Build. Environ.* 169 (2020) 106555. <https://doi.org/10.1016/j.buildenv.2019.106555>.
- [18] A.K. Kar, What Affects Usage Satisfaction in Mobile Payments? Modelling User Generated Content to Develop the “Digital Service Usage Satisfaction Model,” *Inf. Syst. Front.* (2020). <https://doi.org/10.1007/s10796-020-10045-0>.
- [19] S. Abbaszadeh, L. Zagreus, D. Lehrer, C. Huizenga, Occupant satisfaction with indoor environmental quality in green buildings, *HB 2006 - Heal. Build. Creat. a Heal. Indoor Environ. People, Proc.* 3 (2006) 365–370.
- [20] G.R. Newsham, B.J. Birt, C. Arsenault, A.J.L. Thompson, J.A. Veitch, S. Mancini, A.D. Galasiu, B.N. Gover, I.A. MacDonald, G.J. Burns, Do green buildings have better indoor environments? New evidence, *Build. Res. Inf.* 41 (2013) 415–434. <https://doi.org/10.1080/09613218.2013.789951>.
- [21] Z. Gou, S.S.Y. Lau, Z. Zhang, A comparison of indoor environmental satisfaction between two green buildings and a conventional building in China, *J. Green Build.* 7 (2012) 89–104. <https://doi.org/10.3992/jgb.7.2.89>.
- [22] S. Ravindu, R. Rameezdeen, J. Zuo, Z. Zhou, R. Chandratilake, Indoor environment quality of green buildings: Case study of an LEED platinum certified factory in a warm humid tropical climate, *Build. Environ.* 84 (2015) 105–113. <https://doi.org/10.1016/j.buildenv.2014.11.001>.
- [23] L.K. Tee, C. Chee, H.H. Mohamed, O.S. Lee, E-clean: A data cleaning framework for patient data, *Proc. - 1st Int. Conf. Informatics Comput. Intell. ICI 2011.* (2011) 63–68. <https://doi.org/10.1109/ICI.2011.21>.
- [24] P.K. Prerna Mishra, Ranjana Rajnish, Sentiment analysis of twitter data: Case study on digital India, (2016) 170–184. <https://doi.org/10.4018/978-1-5225-3787-8.ch011>.
- [25] No title, (n.d.). <https://www.usgbc.org/projects>.

This paper is under review

- [26] W.J. Wilbur, K. Sirotkin, The automatic identification of stop words, *J. Inf. Sci.* 18 (1992) 45–55. <https://doi.org/10.1177/016555159201800106>.
- [27] V.S. Pagolu, K.N.R.C. Challa, G. Panda, B. Majhi, Sentiment analysis of Twitter data for predicting stock market movements, (2016) 1345–1350.
- [28] Y.S. Chen, L.H. Chen, Y. Takama, Proposal of LDA-Based Sentiment Visualization of Hotel Reviews, *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015.* (2016) 687–693. <https://doi.org/10.1109/ICDMW.2015.72>.
- [29] W. Zhao, J.J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, W. Zou, A heuristic approach to determine an appropriate number of topics in topic modeling, *BMC Bioinformatics.* 16 (2015) S8. <https://doi.org/10.1186/1471-2105-16-S13-S8>.
- [30] D. Putthividhya, H.T. Attias, S.S. Nagarajan, Topic regression multi-modal latent dirichlet allocation for image annotation, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2010) 3408–3415. <https://doi.org/10.1109/CVPR.2010.5540000>.
- [31] Q. Chen, L. Yao, J. Yang, Short text classification based on LDA topic model, *ICALIP 2016 - 2016 Int. Conf. Audio, Lang. Image Process. - Proc.* (2017) 749–753. <https://doi.org/10.1109/ICALIP.2016.7846525>.
- [32] D. Ghosh, R. Guha, What are we “tweeting” about obesity? Mapping tweets with topic modeling and Geographic Information System, *Cartogr. Geogr. Inf. Sci.* 40 (2013) 90–102. <https://doi.org/10.1080/15230406.2013.776210>.
- [33] C. Jacobi, W. Van Atteveldt, K. Welbers, Quantitative analysis of large amounts of journalistic texts using topic modelling, *Digit. Journal.* 4 (2016) 89–106. <https://doi.org/10.1080/21670811.2015.1093271>.
- [34] K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring topic coherence over many models and many topics, *EMNLP-CoNLL 2012 - 2012 Jt. Conf. Empir. Methods Nat. Lang. Process. Comput. Nat. Lang. Learn. Proc. Conf.* (2012) 952–961.
- [35] C. Fellbaum, *WordNet*, Wiley Online Library, 1998.
- [36] USGBC, LEED v4 for neighborhood development, 2009 (2014) 71.
- [37] USGBC, LEED v4 for interior design and construction, (2015).
- [38] USGBC, LEED v4 for building operations and maintenance, (2018) 116.
- [39] USGBC, LEED v4 credits for building design and construction, (2019) 147.
- [40] A. Alsaedi, M.Z. Khan, A study on sentiment analysis techniques of Twitter data, *Int. J. Adv. Comput. Sci. Appl.* 10 (2019) 361–374. <https://doi.org/10.14569/ijacsa.2019.0100248>.
- [41] J.A. Morente-Molinera, G. Kou, Y. Peng, C. Torres-Albero, E. Herrera-Viedma, Analysing discussions in social networks using group decision making methods and sentiment analysis, *Inf. Sci. (Ny).* 447 (2018) 157–168. <https://doi.org/10.1016/j.ins.2018.03.020>.
- [42] A. Bermingham, M. Conway, L. McInerney, N. O’Hare, A.F. Smeaton, Combining social network analysis and sentiment analysis to explore the potential for online radicalisation, *Proc. 2009 Int. Conf. Adv. Soc. Netw. Anal. Mining, ASONAM 2009.* (2009) 231–236. <https://doi.org/10.1109/ASONAM.2009.31>.
- [43] E. Frank, L. Trigg, G. Holmes, I.H. Witten, Technical note: Naive Bayes for regression, *Mach. Learn.* 41 (2000) 5–25. <https://doi.org/10.1023/A:1007670802811>.
- [44] S. Liu, Z. Wang, S. Schiavon, Y. He, M. Luo, H. Zhang, E. Arens, Predicted percentage dissatisfied with vertical temperature gradient, *Energy Build.* 220 (2020). <https://doi.org/10.1016/j.enbuild.2020.110085>.

- [45] Tableau, Tableau Prep Help, Tableau. (2020). https://onlinehelp.tableau.com/current/prep/en-us/prep_welcome.htm.
- [46] D.A. Prum, T. Kobayashi, Green building geography across the United States: Does governmental incentives or economic growth stimulate construction?, SSRN Electron. J. 43 (2013). <https://doi.org/10.2139/ssrn.2276185>.
- [47] L.Y. Heng, R. Logeswaran, B.P. Marikannan, Performance evaluation of analytics models for trends analysis of news, J. Phys. Conf. Ser. 1712 (2020). <https://doi.org/10.1088/1742-6596/1712/1/012021>.
- [48] J. Chang, J. Boyd-Graber, C. Wang, S. Gerrish, D.M. Blei, Reading tea leaves: How humans interpret topic models, (2009).
- [49] USGBC, LEED v4.1 residential BD+C multifamily homes, (2020).
- [50] BREEAM, BREEAM UK new construction, (2018) 634.
- [51] J.H. Scofield, Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings, Energy Build. 67 (2013) 517–524. <https://doi.org/10.1016/j.enbuild.2013.08.032>.
- [52] Y. Roderick, D. Mcewan, C. Wheatley, C. Alonso, Comparison of energy performance assessment between leed, breeam and green star, IBPSA 2009 - Int. Build. Perform. Simul. Assoc. 2009. (2009) 1167–1176.
- [53] USGBC, LEED v4.1 residential BD+C, multifamily homes and multifamily homes core and shell (aparted for India), (2020).
- [54] M.A. Davis, W.D. Larson, S.D. Oliner, J. Shui, The price of residential Land for counties, ZIP codes, and census tracts in the United States, J. Monet. Econ. (2020). <https://doi.org/10.1016/j.jmoneco.2020.12.005>.

Appendix I

Word dictionary

Bins	Topics (Seed words)	Words dictionary
Location & Transportation	Location	school, warehouse, center, retail, healthcare, community, public, mail, surrounding, farmland, industry, residential, site, visitability, alleys, safe, security stores, hospital, landscape, landmark, courtyard, plaza, grocery, market, supermarket, bank, mall, theater, pharmacy, gym, laundry, library, clinic, university, restaurant, warehouse, hotel, vendor, Church, club, studio, café, college, healthcare, telecommunication, landfill, waterway, housing, daycare, education, postsecondary, nursery, sun, winds, weather, rain, mapping, slope, stability, flood, wetlands, lakes, streams, shorelines, rainwater;
	Transportation	walking, vehicle, bus, stop, station, bicycle, Uber, car, ferry, sidewalk, pedestrian, transit, corridor, distance, commute, carpool, connectivity, shuttles, pavement, walkways, roadways, bikeway, path, convenience, motor, automobile, streetcar, rail, carshare, rideshare, passenger, route, tour, trips, travel, freight, lane, on-street, off-street, fleet, truck, dock, conveyance;
Running cost	Cost,	Bill, money, capital, rent, payment, charge, maintenance, purchase, price, economic, sustainable, consumption, efficiency, saving, budget, power, depletion, lifecycle, waste, value, burden, discount, finance, benefits, water, gas, fuel, oil, steam, electricity, propane, load, metering, grid, solar, PV;
Health & Wellbeing	Amenities	pool, park, garden, seating, parking, utility, recreation, entertainment, sports, infrastructure, sanitation, flora, fauna, planting, vegetation, trees, greenfield, view
	Management services	Respond, assist, service, lease, office, management, staff, arrangement, , agency, housekeeping, regulation, code, policy, trash, , stewardship, organize, contractor, , , hospitality,

	cleaning, cleanliness,, pest, leakage, hygiene, recycling, repair;
Pet-policy	Dog, cat, pet;
Appliance	equipment, machine, elevator, fryer, griddle, drawer, cooker, cooking, toaster, refrigerator, freezer, device, dishwasher, steamer, stove, oven, range, drain, tank, disposer, plug, fireplace, woodstove, showerhead, plumbing, network, broiler, faucet, lamp, alarm;
Indoor environment	Thermal comfort: hot, warm, heat, cold, cool, chill, temperature, radiation; Acoustic: noise, voice, sound, loud; Lighting: daylight, luminaire, luminance, illuminance, sunlight, dim; IAQ: CO ₂ , smoke, smoking, fume, odour, contaminant, airflow, emission, VOC, moisture, humidity, ozone, particle, smell, pollution, ventilation, air-conditioning, fan, exhaust, combustion, chemical, biological, particulate, toxicity, vent, formaldehyde, tobacco, exposure; Layout: restroom, ceiling, roof, floor, window, bathroom, design, space, cabinet, locking, hallway, stairwell, closet, basement, porch, lavatory, urinals, balcony, rooftop, toilets, shower, furniture, kitchen, storage, doors, material, garage, wallwash, paints, coating, carpet, area, vestibule, refurbish, gate, decorate, cabinetry, lobby, construction;