The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0959-3845.htm

Understanding host marketing strategies on Airbnb and their impact on listing performance: a text analytics approach

Yeojin Chung
Kookmin University, Seoul, Republic of Korea, and
Surendra Sarnikar
California State University East Bay, Hayward, California, USA

Received 20 October 2020
Revised 15 June 2021
14 August 2021
Accepted 30 September 2021

Abstract

Purpose – Peer-to-peer (P2P) accommodation sharing has become a significant part of the travel and lodging industry, allowing homeowners to engage in entrepreneurial activity via sharing of resources. However, there is limited understanding of how hosts can use listing descriptions to better match their offerings to different consumer segments. The purpose of this paper is to understand the use of listing descriptions by Airbnb hosts and the impact of such descriptions on sales performance.

Design/methodology/approach – In this paper, a deep learning-based sentence-level aspect mining approach is used to extract various aspects from host-provided listing descriptions. Then a regression-based approach is used to understand the impact of various aspects of listing descriptions on listing performance.

Findings – It was found that aspects for which listing descriptions are the sole source of information have the greatest influence on listing performance. The authors also find that the impact of an aspect on listing performance varies by listing type, and that there is a mismatch between the most included aspects by hosts in their listing descriptions and the most influential aspects that impact sales.

Originality/value – The impact of consumer reviews in the context of Airbnb has been extensively studied. A novel aspect of this study is the exploration of P2P accommodations from a supplier perspective, by understanding the use and impact of host-provided textual descriptions on sales. The findings of this study can help better market properties from a practice perspective and better understand consumer information consumption from a theoretical perspective. The authors also demonstrate a new approach for exploring social phenomena by performing quantitative analysis on textual data using deep-learning and regression-based techniques.

Keywords Sharing economy, Airbnb, Host marketing, Text analytics, Deep learning, Aspect mining, Listing performance

Paper type Research paper

1. Introduction

The sharing economy is a major growing segment of the global economy. In the USA, 72% of American adults have used at least one of multiple forms of shared or on demand services (Smith, 2016). The impact of sharing economy is even higher in economies such as China where sharing economy providers registered a transaction volume of USD 473 billion in 2019 and is expected to expand between 10 and 15% annually between 2020 and 2022 ("China's sharing economy: report," 2020). The growth of the sharing economy has given rise to a new generation of entrepreneurs who rely on sharing economy platforms for income generation (Manyika *et al.*, 2016).

Among the large companies offering platforms for the sharing economy is Airbnb, which allows homeowners to engage in entrepreneurial activity by renting their homes to different consumer segments. For homeowners or hosts to attract customers and generate revenue while

The authors would like to acknowledge "R&D Program for Forest Science Technology (Project No. 2019150B10-2123-0301)" provided by Korea Forest Service (Korea Forestry Promotion Institute) for the support.

Information Technology & People

0959-3845
DOI 10.1108/ITP-10-2020-0718

meeting the constraints of their offerings, it is important for hosts to know how to market their offerings to improve sales and appropriately target customer segments. Such marketing activity in sharing economy platforms are very different from traditional marketing approaches (Eckhardt *et al.*, 2019) in that a majority of the marketing activity is through the sharing economy platform and is carried out by entrepreneurs with limited marketing resources.

A key attribute of product information on Airbnb is the listing descriptions provided by hosts. In the absence of brand information and other traditional marketing cues, the listing description is an important source of listing information for customers and an equally important mechanism for hosts to market their offerings. Listing description can provide information on facility, neighborhood and other aspects of the product that cannot be communicated visually or quantitatively. The listing description can also be used to specifically target specific consumer segments. Given the potential uses of listing description from a marketing perspective, we intend to first study how property owners are using listing descriptions to market their properties, leading to our first research question.

RQ1. What are the common aspects that marketers focus on when marketing their property in the sharing economy?

There are several studies that explore the impact of Airbnb customer reviews and its impact on customer satisfaction and listing performance. However, our understanding of the impact of listing descriptions on listing performance is limited. Past research indicates that product descriptions have a significant impact on sales, especially in the context of e-commerce and experiential products. Given the importance of listing descriptions for hosts or entrepreneurs as a key marketing tool, studying the effect of listing descriptions on sales performance can help improve our understanding of marketing strategies in the sharing economy and help entrepreneurs better market their offerings. Moreover, research by Lutz and Newlands (2018) in the context of Airbnb suggests the presence of various customer segments within the market and that hosts may benefit from customized marketing to such customer segments. However, it is unclear the extent to which such marketing logic is applied by hosts. In order to further explore this issue and understand how hosts' provided listing descriptions impact listing performance, we pursue the following two research questions:

RQ2. How do different aspects in marketer-generated content influence listing performance for shared accommodations?

RQ3. Does the impact of marketer-generated content on listing performance vary based on market segments?

In this paper, we use text mining and regression-based approaches to study host provided listing descriptions and identify its impact on listing performance. Specifically, we first use a deep learning-based approach to extract aspects from host-provided listing descriptions. We then use regression analysis to study the effect of various aspects of listing description on listing performance and their varying impacts in three different market segments. The rest of the paper is structured as follows: We present an overview of relevant literature in this area and our research objectives in Section 2. We then present details of our dataset and research method, including the deep learning-based text mining approach and our statistical model in Section 3, followed by the results and analysis in Section 4. We discuss our findings and its implications for practice and research in Section 5 and conclude with a summary of our contributions in Section 6.

2. Literature review

2.1 Consumer information search

Product uncertainty is a major concern in online markets where users cannot easily evaluate products. This is especially the case with experience goods such as accommodations

(Dimoka *et al.*, 2012). Consumers reduce such uncertainty by using multiple sources of information including user-generated content (UGC) such as customer reviews or marketer-generated content (MGC) such as product descriptions. In the context of sharing economy platforms such as Airbnb, UGC including customer reviews have been found to have an impact on listing performance (Li *et al.*, 2019; Zhang, 2019a). However, the impact of MGC such as listing descriptions is not well researched.

In Airbnb, marketers use listing descriptions to emphasize various aspects of their property (Liang *et al.*, 2020). The impact of various aspects in the listing description can be studied using the economics of information theory (Stigler, 1961) and its extensions in the hospitality industry (Gursoy, 2019; Gursoy and McCleary, 2004). Consumers utilize multiple information sources to make purchase decisions based on cost and benefit of information. Costs include cognitive processing costs and can be different for different consumer segments. The use of an information source is influenced by "perceived cost of information search, level of cognitive processing required, level of consumer processing capabilities, and level of consumer involvement with product" (Gursoy, 2019). Listing descriptions are likely to be lowest cost information sources with most benefits for aspects that cannot be easily found using customer reviews or third-party sources. Such aspects, where listing descriptions are the sole or primary source of information are likely to be significantly associated with listing performance. The information costs are also likely to vary based on consumer segments. Consumers looking for longer stays are likely to be more involved and invest more in information search compared to those looking for short stays. Therefore, an aspect's importance and impact on listing performance is also likely to vary by market segment.

In the next section, we discuss relevant research on aspects extraction in the context of Airbnb and related insights into consumer expectations, followed by a review of extant research on factors affecting listing performance in Section 2.3.

2.2 Consumer expectations in Airbnb

There is a significant amount of literature that has studied the sharing economy phenomenon in the context of Airbnb (Dann *et al.*, 2019), including many that have utilized text mining-based approaches to study customer reviews and UGC (Cheng and Jin, 2019; Lawani *et al.*, 2019; Lee *et al.*, 2019; Luo and Tang, 2019; Zhang, 2019a; Zhang *et al.*, 2018). Consumer expectations in sharing economy vary significantly from traditional businesses. For example, in a comparative study of customer online reviews between traditional hotels and Airbnb accommodations, Zhang (2019b) finds significant differences in review topics, and multiple unique topics in Airbnb reviews including "food in kitchen", "help from host" which are not present in traditional hotel reviews. Lee *et al.* (2019) analyzed customer reviews to understand aspects that influence customer experience. They find that amenities, cleanliness, homeliness, host attributes, location and transportation were common aspects mentioned in reviews across the years. Location, amenities and host were also identified as key aspects in a study of Airbnb customer reviews by Cheng and Jin (2019).

In addition to customer reviews, past studies have also explored the impact of MGC such as host profiles and their impact on customer trust. Tussyadiah and Park (2018) explore host profile descriptions to understand the impact of host self-presentation strategies and its implications on trust. According to their study, consumers show higher trust in hosts who come across as well-traveled hosts rather than those with professional profiles. Zhang *et al.* (2018) also find that host self-descriptions that focus on service, interaction and familiarity with location rather than profession and personality are seen as more trustworthy by consumers.

In order to successfully market their listings in a sharing economy platform, hosts need to tailor their listing descriptions to address the expectations of the sharing economy consumer. While past studies on customer reviews and host profiles provide an insight into customer expectations, the extent to which such expectations are reflected in listing descriptions is not known.

2.3 Factors affecting listing performance

Several studies have studied factors affecting listing performance in the context of Airbnb. Xie and Mao (2017) find that host quality and quantity attributes have a significant impact on listing performance through cue base trust. Specifically, host attributes such as being a superhost, experience as an Airbnb host and host response rate were positively related to number of reservations. Biswas *et al.* (2020) study marketer- and user-generated features and find that among host-generated features, superhost status, experience as a host and amenities are significant predictors of listing performance. Nieto García *et al.* (2019) study host self-presentation strategies and its impact on revenues. They find that revenues are higher when host self-presentation content focusses on social values.

Most studies that explore factors impacting sales or listing performance have focused on host attributes (Biswas *et al.*, 2020; Nieto García *et al.*, 2019; Xie and Mao, 2017) or customer reviews (Li *et al.*, 2019; Zhang, 2019a). However, past research has shown that in the context of lodging websites in the sharing economy, lodging information has a positive effect on customer purchase intentions (Nisar *et al.*, 2019). Listing descriptions therefore are a key marketing tool for hosts and can significantly influence listing performance (Liang *et al.*, 2020). In their study on the impact of listing descriptions, Liang *et al.* (2020) find that the breadth and depth of listing descriptions have a positive impact on listing performance. In this paper, we further study listing descriptions by using text-mining approaches to identify the aspects that are present in listing descriptions and estimating their impact on listing performance in different market segments. A summary of the key literature in this area, and the specific research question addressed by the current paper in the context of relevant previous research is presented in Table 1.

3. Methods

3.1 Data collection

In order to address our research questions, we collected data on Airbnb listings from the website *InsideAirbnb.com*. The website provides publicly available information about Airbnb listings from several cities in cleansed and aggregated formats. The data consist of various text descriptions and key metrics about listings and hosts. We collected listing descriptions from the November 2019 snapshot of listings of two adjacent cities, San Francisco and Oakland in the USA for our analysis. In total, 103,616 sentences from 11,845 listings were included in the data for training the aspect extraction model from text description. In order to explore the effect of the current listing descriptions through the review per month of the upcoming quarter, we collected data from October 2018 to December 2019. Aspects of listing descriptions were extracted based on the pretrained model and summarized quarterly for use in a negative binomial mixed effect model. The final dataset for negative binomial regression analysis includes 14,506 observations during four-quarters and 5,473 listing IDs.

We combined the data from Airbnb with other economic and geographic data obtained from the American Community Survey and geographic information system (GIS) datasets. We retrieved socioeconomic data such as percentage of residents with bachelor's degree or above and per capita income from the 2018 American Community Survey. For each listing, we identified the relevant census tract using the latitude and longitude co-ordinates of the listing locations and identified the relevant socioeconomic variables. The distance of each location to the central business district was determined using shapefiles for each city's central business district, and using *qGIS,* we measured the distance to the centroid of the central business district.

Reference	Study focus	Findings
Tussyadiah and Park (2018)	Focus: host self-presentation	Consumers show higher trust in hosts who display a "well-traveled" profile than "professional" profile
Xie and Mao (2017)	Focus: host attributes	Host quality attributes (local host, superhost, responsiveness, verification, experience) have a positive impact on listing performance
Zhang (2019a)	Focus: customer preferences	Customer reviews indicate preferences for clean properties that provide a home-like feeling and quick response from the host
Biswas *et al.* (2020)	Focus: host attributes	Among marketer-generated content, superhost status, host duration, bedrooms and amenities were significant predictors of listing performance
Liang *et al.* (2020)	Focus: listing descriptions	Providing comprehensive and detailed descriptions of property improves review volume
Nieto García *et al.* (2019)	Focus: host self-presentation (limited to private rooms)	Revenues are higher when host self-presentation content focusses on social values
Lee *et al.* (2019)	Focus: consumer preferences	Amenities, cleanliness, homeliness, host attributes, location and transportation were common aspects mentioned in reviews across the years. Some aspects were seasonal such as access to swimming pools during summer or hot water for showers during winter
Cheng and Jin (2019)	Focus: consumer preferences	Most common aspects of consumer reviews were location, amenities and host. Users assess their stay experiences using attributes similar to hotel stays
Zhang *et al.* (2018)	Focus: host self-descriptions	Host self-descriptions that focus on interaction, services and familiarity with location rather than profession and personality are seen as more trustworthy
This study	*Focus: listing descriptions*	*RQ: What aspects are typically present in listing descriptions? How do they impact listing performance in different markets?*

Table 1. Summary of literature review

3.2 Text preprocessing

We first tokenized the listing descriptions into words. Second, we removed non-English words and special characters including punctuations (e.g. "*", ":", "–"). We also filtered out stop words, which are commonly used but not informative in aspect mining (e.g. "I", "a", "and", "do"). Third, the words were lemmatized to transform inflectional forms and derivationally related forms of a word to a common base form (e.g. "playing", "played", "plays" → "play"). Finally, we retained only nouns, proper nouns, adjectives, verbs and adverbs in the data, and then built the corpus with the most frequent 2,000 words for analysis. In order to reduce ambiguity of aspects, we removed short sentences of size two words or lower during preprocessing. The python package *nltk* (Bird *et al.*, 2009) was used for preprocessing steps.

3.3 Aspect-mining methods

Aspect extraction has been studied extensively in past decades, and it has played an important role in natural language processing. Aspect extraction identifies aspect terms from text data and clusters them into similar aspects. For example, in the sentence "There are all the basics needed for cooking", an aspect word "cooking" is identified and grouped into aspect "kitchen" with other aspect words such as "spice" and "dish". Most of the previous works on aspect extraction can be categorized into supervised learning and unsupervised learning approaches.

Supervised learning approaches regard aspect extraction as a standard sequence labeling problem (Jakob and Gurevych, 2010; Jin and Ho, 2009). With recent development of deep learning methodology, several supervised methods using deep neural networks, convolutional neural networks and recurrent neural networks have been proposed (Wang *et al.*, 2016; Zhang *et al.*, 2015). While supervised deep learning-based aspect extraction provides good performance, it requires a large amount of labeled data for training purpose, and it needs to be domain specific (Araque *et al.*, 2017; Chen *et al.*, 2017). A major limitation of supervised approaches from a research perspective is that the aspects need to be predetermined and provided by the researcher as opposed to emerging from the dataset.

Unsupervised learning approaches have been proposed in many applications where a significant amount of labeled data is not available, and the need is to explore and model the aspects that emerge from the data. In this case, aspect extraction is often achieved by the topic modeling approach. Majority of existing works are variants of the latent dirichlet allocation (LDA) model (Blei *et al.*, 2003; Minka and Lafferty, 2002) with modifications to incorporate document-level covariates (Blei and McAuliffe, 2007; Mimno and McCallum, 2008; Roberts *et al.*, 2016). Despite its popularity, the induced aspects from the LDA-based models were often found to be of poor quality due to the fact that the LDA-based models do not directly model the word co-occurrence and do not estimate the topic distributions efficiently when the documents tend to be short (He *et al.*, 2017; Luo *et al.*, 2019b; Mimno *et al.*, 2011). To this end, deep learning-based models have recently been proposed for aspect extraction and have shown significantly improved performances (He *et al.*, 2017; Luo *et al.*, 2019a; Wang *et al.*, 2015).

The *attention-based aspect extraction* (ABAE; He *et al.*, 2017) model is a deep learning-based unsupervised model to incorporate the attention mechanism to focus more on aspect words than others which are not related to any aspects. Unlike LDA-based models, the ABAE model considers word co-occurrence directly in word embeddings and learns aspect embeddings in the same embedding space with word embedding. For each aspect, representative words are chosen by its closest words based on the cosine similarity in the embedding space. The ABAE model is shown to produce more meaningful aspects and performs better than LDA-based models for sentence-level aspect extraction.

In the following analysis, we use the ABAE method to extract aspects from listing descriptions. Since a typical listing description consists of multiple sentences and aspects, we split the description into sentences and find aspects for each sentence. The ABAE model is particularly well suited for and shown to perform well with sentence-level aspect extraction.

3.4 Data analysis

We use a two-stage approach to analyze the data. We begin with the ABAE method to extract aspects from the descriptions and identify common aspects emphasized by the hosts in their listing descriptions. We then use a negative binomial mixed effect model to estimate the effect of the aspects on listing performance.

3.4.1 Aspect extraction. The ABAE model takes the word embedding vectors as input data. We calculated the embeddings using the *skip-gram* (Mikolov *et al.*, 2013) method with the dimension 200, following He *et al.* (2017). The number of aspects from 8 to 12 were considered. Each model was learned with 100 epochs and a batch-size of 50.

The optimal number of aspects was chosen based on *topic coherence scores* (Mimno *et al.*, 2011). The topic coherence score is known to correlate highly with human judgment and has been regarded as a standard method to measure the quality of topic models (Roberts *et al.*, 2016; Röder *et al.*, 2015). Given an aspect z and a set of the N most probable words for aspect z, $S^z = \{w_1^z, w_2^z, \ldots, w_N^z\}$, the coherence score is defined as below:

$$C(z;\, S^z) = \sum_{m=2}^{N} \sum_{l=1}^{n-1} \log \frac{D\left(w_m^z,\, w_l^z\right) + 1}{D\left(w_l^z\right)}$$

where $D(w)$ is the number of documents containing word w and $D(w,\, w')$ is the number of documents containing word w and w' at the same time.

Figure 1 compares coherence scores of the ABAE models with different number of aspects. For each value of N $(= 10,\, 20,\, \ldots,\, 60)$, the coherence score of ten aspects is the largest or close to the largest among the number of aspects considered. Therefore, we select the ABAE model with ten aspects and detail the results based on this model in Section 4.1.

We used an iterative process for naming the aspects. First, we analyzed aspect extraction results from other research on Airbnb (Cheng and Jin, 2019; Zhang, 2019a) and used similar naming conventions (for e.g. room, facilities, communication, kitchen, etc.) for aspects where there was a significant overlap in keywords and concepts. Each researcher then independently labeled each of the remaining aspects. The name for an aspect was selected when both researchers agreed on the name followed by a discussion on labels where there was a disagreement. The process was repeated until both researchers arrived at a consensus on the naming of the aspects. However, given the subjectivity involved in the naming process, we also provide the extracted list of keywords for each aspect, so the aspects can be compared with other studies based on keyword overlaps.

3.4.2 Performance analysis of the listings. Following aspect extraction, which provides us with the most commonly used aspects in listing descriptions, we employ a negative binomial mixed effect model to estimate the influence of marketing strategy of a host as defined by their focus on various aspects in the listing description on the performance of the listing. Several researches have used the number of reviews as proxy for listing performance (Liu *et al.*, 2018; Ye *et al.*, 2011; Zhang, 2019a). Therefore, we calculated the number of reviews for listing i in the quarter t and included it as a response variable in the model.

For independent variables, we measured marketing strategy of a host with frequencies of the aspects mentioned in the listing description. For listing i in quarter t, the number of sentences assigned to each aspect was counted to construct ten aspect variables

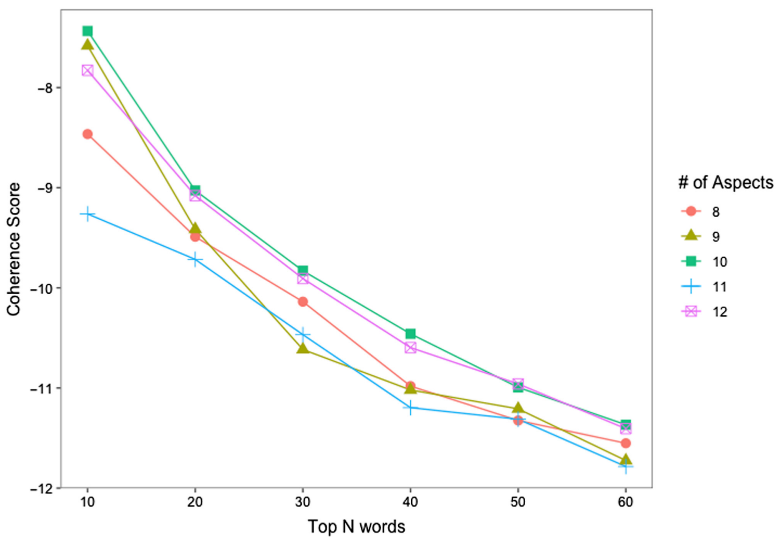

Figure 1.
Coherence scores

($Comm_{it}$, $Attr_{it}$, . . .). If a listing description was changed within the quarter, the most recent one was kept in the data. In order to compare the performance between listing types, we used two dummy variables ($D_{PrivateRoom,i}$, $D_{SharedRoom,i}$). The models also include the distance to downtown ($Downtown_i$), per capita income in the surrounding area ($PerCapitaIncome_i$) and a dummy variable for the city of San Francisco ($D_{SanFrancisco,i}$) as control variables. Customer reviews have been known to influence the performance significantly in the previous studies (Rosario *et al.*, 2016; Zhang, 2019a). We calculated average review scores over three months within a quarter ($Review\_scores\_rating\_avg_{it}$) and included it as a control variable. We also included in our model two host attributes: a dummy variable for "Superhost" ($D_{Superhost}$) and the duration of host's activity on Airbnb ($Host\_duration$). Airbnb "Superhost" status is given to the hosts providing outstanding services, and recent studies have shown that the "Superhost" badge has a positive impact on the volume of customer reviews (Biswas *et al.*, 2020; Chattopadhyay and Mitra, 2019; Ert and Fleischer, 2019; Gunter, 2018; Liang *et al.*, 2017) The host duration has also been regarded as a determinant of the number of customer reviews in the previous studies (Biswas *et al.*, 2020; Xie and Mao, 2017; Xu, 2020). All the quantitative predictors were standardized.

We used negative binomial mixed effect model to analyze the data. Since our response variable (the number of reviews in a quarter) is an observed count and its variance (49.96) is much larger than its mean (7.73), it is appropriate to regard our dependent variable as observations from a negative binomial distribution. Furthermore, to account for within-listing correlations across four-quarters, we include random intercept term in the model. In the negative binomial mixed effect model, the count y_{it} are assumed to follow the negative binomial distribution:

$$y_{it} \sim NB(y_{it}|\mu_{it}, \theta) = \frac{\Gamma(y_{it}+\theta)}{\Gamma(\theta)y_{it}!} \cdot \left(\frac{\theta}{\mu_{it}+\theta}\right)^{\theta} \cdot \left(\frac{\mu_{it}}{\mu_{it}+\theta}\right)^{y_{it}}$$

where μ_{it} and θ are the mean and the dispersion parameters, respectively. The mean parameter μ_{it} for listing i in quarter t is modeled via the link function logarithm.

To study the differential impacts of aspects on listing performance by listing types, we model the interaction effects between listing types and aspect variables. For this purpose, we tested four different models. Model 1 includes all the independent variables listed above. For Model 2, six control variables are added to Model 1. Model 3 contains the 20 interaction terms between the two listing type dummies and the ten aspect variables. We then use the backwards elimination method to build a more parsimonious Model 4 by dropping the least significant predictor sequentially with $\alpha = 0.01$. The four models are defined as follows:

[Model 1]

$$\log(\mu_{it}) = \beta_0 + \beta_1 D_{PrivateRoom,i} + \beta_2 D_{SharedRoom,i} + \beta_3 Attr_{i,t-1} + \ldots + \beta_{12} Facil_{i,t-1}$$

[Model 2]

$$\log(\mu_{it}) = \beta_0 + \beta_1 D_{PrivateRoom,i} + \beta_2 D_{SharedRoom,i} + \beta_3 Attr_{i,t-1} + \ldots + \beta_{12} Facil_{i,t-1} \\ +\beta_{13} Downtown_i + \ldots + \beta_{17} Host\_duration_{i,t-1} + \beta_{18} D_{Superhost,i,t-1}$$

[Model 3]

$$\log(\mu_{it}) = \beta_0 + \beta_1 D_{PrivateRoom,i} + \beta_2 D_{SharedRoom,i} + \beta_3 Attr_{i,t-1} + \ldots + \beta_{12} Facil_{i,t-1} \\ +\beta_{13} Downtown_i + \ldots + \beta_{17} Host\_duration_{i,t-1} + \beta_{18} D_{Superhost,i,t-1} \\ +\beta_{19} D_{PrivateRoom,i} \cdot Attr_{i,t-1} + \ldots + \beta_{38} D_{SharedRoom,i} \cdot Facil_{i,t-1}$$

[Model 4]

$$\log(\mu_{it}) = \beta_0 + \beta_1 D_{PrivateRoom,i} + \beta_2 D_{SharedRoom,i} + \beta_3 Attr_{i,t-1} + \ldots + \beta_{12} Facil_{i,t-1} + \beta_{13} Downtown_i + \ldots \beta_{17} Host\_duration_{i,t-1} + \beta_{18} D_{Superhost,i,t-1} + \beta_{17} D_{PrivateRoom,i} \cdot Bldg_{i,t-1} + \beta_{18} D_{PrivateRoom,i} \cdot Style_{i,t-1} + \beta_{19} D_{PrivateRoom,i} \cdot Kitchen_{i,t-1} + \beta_{20} D_{SharedRoom,i} \cdot Comm_{i,t-1} + \beta_{21} D_{SharedRoom,i} \cdot Bldg_{i,t-1} + \beta_{22} D_{SharedRoom,i} \cdot Visitor_{i,t-1}$$

4. Results

4.1 Aspect extraction and exploratory analysis

Table 2 presents representative words and top sentences for the ten inferred aspects extracted from the ABAE model. Figure 2 displays the total number of sentences in each aspect. The most frequently appearing aspect is "attraction/transportation" (12.9%), indicating attractions near the property and transportation. The next most frequent aspects are as follows: "communication" (12.9%, how to communicate with the host), "building structure" (12.2%, location of the rooms and facilities on the building structure), "visitor" (10.0%, typical types of visitors), "room" (9.9%, details of the bedroom), "interior style" (9.6%, interior design style and renovation), "nature" (8.9%, natural surroundings), "neighborhood" (8.6%, characteristics of neighborhood), "kitchen" (7.9%, kitchen supplies and equipment) and "facility" (7.2%, indoor and outdoor facilities at the property). In Figure 2, we observe that there were much fewer sentences for shared rooms than for entire homes or private rooms regardless of the aspects. This is caused by the fact that the dataset contains only 195 shared room listings, while it has 3,240 for entire homes and 2,038 private room listings.

To compare the distribution of aspects by listing types, we calculated the average number of sentences inferred to each aspect, shown in Figure 3. The descriptions of entire homes tend to include more sentences about "interior style", "nature", and "neighborhood" than the other types of listings. The private rooms share the similar shape of the chart with entire homes but tend to have more sentences about "communication", "building structure" and "visitor" on average. The shared rooms have clearly a different shape of the chart from the other two. The hosts of the shared rooms are likely to describe much more often about "visitor" and "room" than those of the entire homes and private rooms.

Table 5 contains the top three aspects with the highest average number of sentences for each listing type. We find that "communication" is among the most frequent across all listing types. Entire homes and private room segments both share "attraction/transportation" and "building structure" as a frequently mentioned aspect, whereas shared room descriptions have "visitor" and "room" aspects among the most frequent.

4.2 Aspect influence analysis

A summary of the regression results is presented in Table 3. Models 1 and 2 reveal that the coefficient estimates of the independent variables are not significantly different between the two models since the estimates are located within two standard errors of the corresponding estimates in the other model. Since Model 3 is significantly different from Model 2 based on the likelihood ratio test, we infer that the impact of an aspect could vary depending on the listing type. Comparing Model 3 and Model 4, while there is no statistically significant difference between them, Model 4 has the smallest Bayesian information criterion (BIC) and Akaike information criterion (AIC). Therefore, we use the parsimonious model, Model 4, as our final model.

A significant interaction term confirms that the slope of the aspect variable is different between entire homes and the corresponding listing types. When a significant interaction effects exists, the regression coefficients of the main effects need to be interpreted carefully.

Inferred aspects	Representative words	Top sentences
Communication (Comm)	Send, contact, otherwise, aware, calendar, reserve, regard, availability, arrange, advance, prior, however, schedule, policy, therefore, please, due, emergency, inquire, vary, inquiry, interested, important, taxis, message, remain, sister, reservation, generally, ask	- Please contact for availability/ pricing - Please inquire to reserve - Just send a text
Attraction/ transportation (Attr)	Ride, ferry, symphony, opera, theatre, hospital, mall, fisherman, fox, civic, wharf, waterfront, underground, convention, miles, pier, mile, stop, major, stations, jazz, subway, college, quick, less, route, oracle, lake, attraction, medical	- 30 min BART ride to San Francisco' – It's a 5 min Lyft ride to 19th St - The SF Jazz Center, Ballet, Opera and Symphony are within four blocks
Building structure (Bldg)	Downstairs, exclusive, upstairs, split, adjacent, attach, separate, rear, level, foyer, share, hallway, basement, consist, main, private, bedroom, den, shared, backyard, quarter, duplex, second, entrance, staircase, detach, master, story, upper, total	- Two bedrooms downstairs - We have 2 bathrooms, one upstairs and one downstairs - Your bedroom will be upstairs
Visitor	Visit, adventure, want, together, visitor, entertain, traveler, friend, trip, leisure, holiday, memorable, social, fun, student, pleasure, goal, vacation, base, conversation, explore, getaway, travel, budget, dream, encourage, folk, ability, hope, opportunity	- Perfect for visiting professors or when visiting with students and family - Solo travelers only - Perfect for business travelers
Interior style (Style)	Architectural, contemporary, century, interior, restore, retain, blend, original, vintage, charm, tasteful, touch, detail, industrial, period, classic, sleek, architect, pair, construct, upgrade, finish, era, originally, furnishing, mid, accent, update, decorate, meticulously	- Victorian/Edwardian period details - The interiors were recently redone - Architectural style from 1930
Room	Needle, firm, beds, foam, clock, bed, pillow, king, pull, fold, daybed, radio, memory, couch, plush, armoire, sofa, dresser, leather, lamp, mattresses, sized, size, alarm, deluxe, trundle, convertible, blackout, twin, feather	- Queen bed with Tuft and Needle mattress - Both mattresses are firm - Beds all have down pillows and down comforters
Nature	Bird, fountain, flower, afternoon, sip, giant, glass, herb, hammock, bench, vegetable, fruit, lemon, chicken, swing, listen, wind, morning, pond, lawn, kick, deer, wine, stain, picnic, plant, sun, wild, fog, round	- Figs, lemons, tomatoes, herbs, lettuce, cucumbers, and persimmon - Lots of birds singing in the morning - On the deck is a bird feeder and a bird bath which attract lots of birds
Neighborhood (Nbhd)	Village, quaint, yet, quiet, peaceful, quite, nestled, height, residential, heights, scenic, diamond, lively, triangle, desirable, neighborhood, tuck, corona, bustle, middle, bustling, cute, extremely, neighborhoods, serene, canyon, glen, avenue, safe, road	- The neighborhood is very peaceful and quiet - Quiet and serene' – It's quite residential
Facility (Facil)	Pay, facility, cleaning, unlimited, indoor, rooftop, package, parking, laundry, service, privilege, fitness, valet, pool, gym, security, site, concierge, exercise, courtyard, dry, services, fee, lobby, garage, fire, trash, grill, doorman, elevator	- Laundry facilities onsite - Parking on site - Offsite Valet Parpprox.pprox $35/ day-paid by the guest
Kitchen	Spice, supply, oil, dish, pans, olive, pot, pantry, salt, plate, basic, freezer, utensils, press, cooking, pan, necessity, kettle, prep, maker, shampoo, item, refrigerator, snack, induction, rice, cream, sugar, conditioner, ice	- Coffee is supplied - Kitchen and other supplies - The kitchen is well stocked with all the pots and pans, utensils, spices, etc

Table 2.
Representative words for inferred aspects

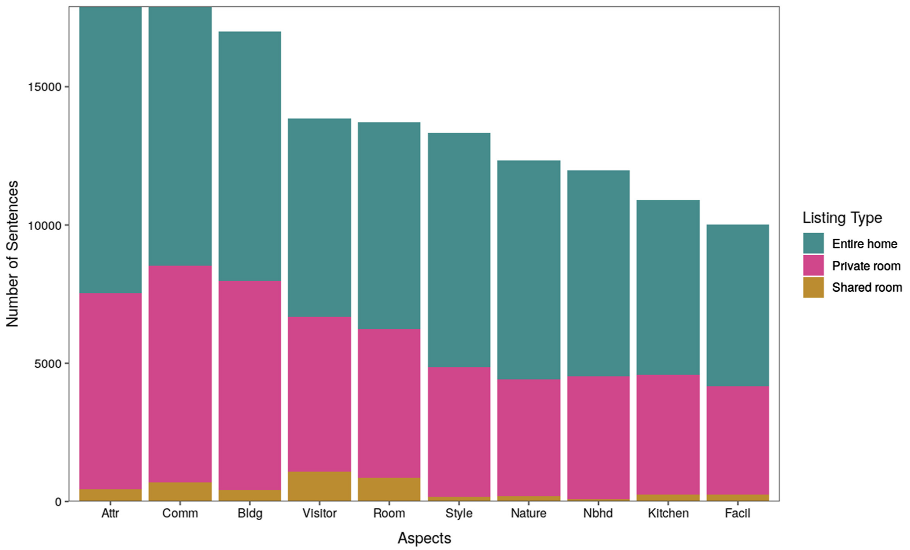

Figure 2.
The number of sentences in aspects

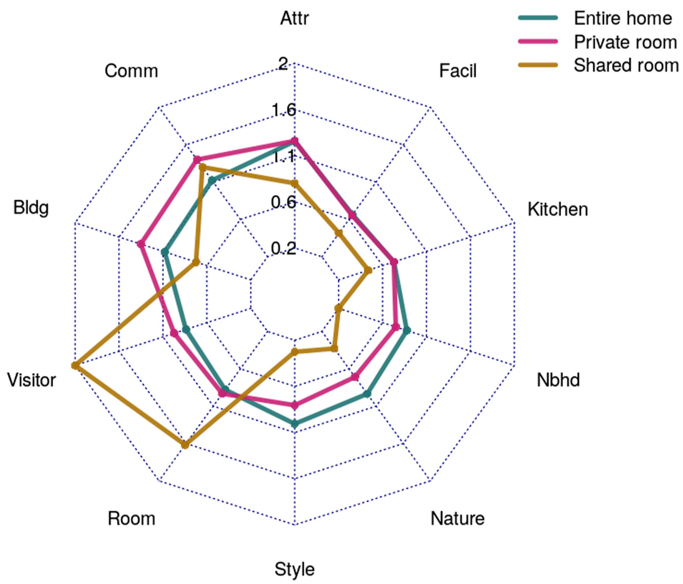

Figure 3.
Comparison of average number of sentences in a listing description by listing types

Therefore, we calculated the slopes of aspect variables for each listing type and compared them. For example, with the significant interaction of $D_{PrivateRoom}$ and *Kitchen*, the slope of *Kitchen* for the private rooms is −0.006: the sum of the slope of *Kitchen* (0.077) and that of $D_{PrivateRoom}$: *Kitchen* (−0.083). Since the interaction between $D_{SharedRoom}$ and *Kitchen* is not in the model, the slopes for the shared rooms and the entire homes are not significantly different, having 0.077. The significance tests for individual slopes were conducted by the multiple comparison method in Hothorn *et al.* (2008).

Table 3.
Coefficient estimates and model comparison

Fixed effect	Model1 Estimate	Model1 Std. error	Model2 Estimate	Model2 Std. error	Model3 Estimate	Model3 Std. error	Model4 Estimate	Model4 Std. error
(Intercept)	1.299	(0.031)***	1.334	(0.046)***	1.346	(0.051)***	1.350	(0.049)***
$D_{PrivateRoom}$	0.212	(0.026)***	0.208	(0.025)***	0.183	(0.057)**	0.166	(0.042)***
$D_{SharedRoom}$	0.152	(0.070)*	0.267	(0.070)***	0.413	(0.137)**	0.450	(0.096)***
Attr	−0.004	(0.008)	−0.001	(0.008)	−0.026	(0.010)**	−0.025	(0.010)*
Comm	0.003	(0.007)	0.004	(0.007)	0.008	(0.009)	0.009	(0.007)
Bldg	0.041	(0.008)***	0.031	(0.008)***	0.052	(0.011)***	0.053	(0.010)***
Visitor	0.015	(0.009)	0.014	(0.009)	0.009	(0.013)	0.009	(0.010)
Room	0.019	(0.010)*	0.013	(0.009)	0.002	(0.012)	0.018	(0.009)
Style	−0.024	(0.010)*	−0.015	(0.010)	−0.042	(0.012)***	−0.041	(0.012)***
Nature	0.017	(0.010)	0.006	(0.010)	0.010	(0.012)	0.010	(0.010)
Nbhd	−0.006	(0.011)	−0.017	(0.011)	−0.004	(0.013)	−0.016	(0.011)
Kitchen	0.047	(0.011)***	0.043	(0.011)***	0.078	(0.015)***	0.077	(0.014)***
Facil	−0.006	(0.010)	−0.002	(0.010)	0.008	(0.013)	−0.004	(0.010)
Downtown			0.138	(0.014)***	0.140	(0.014)***	0.141	(0.014)***
PerCapitaIncome			−0.055	(0.015)***	−0.053	(0.015)***	−0.052	(0.015)***
$D_{SanFrancisco}$			−0.037	(0.032)	−0.042	(0.032)	−0.041	(0.032)
Review_scores_rating_avg			0.086	(0.011)***	0.085	(0.011)***	0.085	(0.011)***
Host_duration			−0.007	(0.005)	−0.009	(0.005)	−0.009	(0.005)
$D_{Superhost}$			0.109	(0.013)***	0.111	(0.013)***	0.110	(0.013)***
$D_{PrivateRoom}$:*Attr*					0.069	(0.015)***	0.064	(0.015)***
$D_{PrivateRoom}$:*Comm*					0.000	(0.014)		
$D_{PrivateRoom}$:*Bldg*					−0.037	(0.015)*	−0.038	(0.015)*
$D_{PrivateRoom}$:*Visitor*					0.001	(0.018)		
$D_{PrivateRoom}$:*Room*					0.033	(0.019)		
$D_{PrivateRoom}$:*Style*					0.079	(0.020)***	0.081	(0.019)***
$D_{PrivateRoom}$:*Nature*					−0.006	(0.020)		
$D_{PrivateRoom}$:*Nbhd*					−0.030	(0.021)		
$D_{PrivateRoom}$:*Kitchen*					−0.083	(0.021)***	−0.083	(0.020)***
$D_{PrivateRoom}$:*Facil*					−0.032	(0.020)		
$D_{SharedRoom}$:*Attr*					−0.063	(0.056)		
$D_{SharedRoom}$:*Comm*					−0.095	(0.041)*	−0.079	(0.037)*

(*continued*)

Table 3.

Fixed effect	Model1 Estimate	Model1 Std. error	Model2 Estimate	Model2 Std. error	Model3 Estimate	Model3 Std. error	Model4 Estimate	Model4 Std. error
$D_{SharedRoom}$:*Bldg*					−0.248	(0.067)***	−0.230	(0.060)***
$D_{SharedRoom}$:*Visitor*					0.064	(0.033)	0.059	(0.028)*
$D_{SharedRoom}$:*Bedroom*					0.051	(0.046)		
$D_{SharedRoom}$:*Style*					0.108	(0.081)		
$D_{SharedRoom}$:*Nature*					0.081	(0.061)		
$D_{SharedRoom}$:*Nbhd*					0.004	(0.097)		
$D_{SharedRoom}$:*Kitchen*					0.000	(0.058)		
$D_{SharedRoom}$:*Facil*					−0.039	(0.075)		
Random effect								
Variance (group: ID)	0.850		0.792		0.777		0.805	
Model comparison								
Likelihood ratio test (compared with the previous model)			294.16	***	99.87	***	12.88	
AIC	78,879		78,597		78,537		78,524	
BIC	78,993		78,756		78,848		78,736	

Note(s): ***: p-value < 0.001, **: 0.001 < p-value < 0.01, *: 0.01 < p-value < 0.05

Table 4. Comparison of regression slopes of aspects

	Entire home		Private room		Shared room	
	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.
Attr	−0.025	(0.010)*	0.039	(0.012)***	−0.025	(0.010)*
Comm	0.009	(0.007)	0.009	(0.007)	−0.071	(0.037)
Bldg	0.053	(0.010)***	0.015	(0.012)	−0.178	(0.060)**
Visitor	0.009	(0.010)	0.009	(0.010)	0.068	(0.026)**
Room	0.018	(0.009)	0.018	(0.009)	0.018	(0.009)
Style	−0.041	(0.012)***	0.040	(0.016)*	−0.041	(0.012)***
Nature	0.010	(0.010)	0.010	(0.010)	0.010	(0.010)
Nbhd	−0.016	(0.011)	−0.016	(0.011)	−0.016	(0.011)
Kitchen	0.077	(0.014)***	−0.006	(0.015)	0.077	(0.014)***
Facil	−0.004	(0.010)	−0.004	(0.010)	−0.004	(0.010)

Note(s): ***: p-value < 0.001, **: 0.001 < p-value < 0.01, *: 0.01 < p-value < 0.05

Table 5. Summary of the most influential and most frequent aspects by listing type

Listing type	The most effective aspects	The most frequent aspects
Entire home	Kitchen	Attraction/transportation
	Building structure	Communication
	Room	Building structure
Private room	Interior style	Communication
	Attraction/transportation	Building structure
	Room	Attraction/transportation
Shared room	Kitchen	Visitor
	Visitor	Room
	Room	Communication

Table 4 displays the slopes for the aspect variables, their standard errors and significance in each listing type. For the entire homes, "kitchen" and "building structure" have a significant positive impact on the listing performance. In particular, "kitchen" has the largest slope of 0.077 among them. We observe that the regression lines for entire homes and shared rooms are steeper than that of private room. This implies that more description about the kitchen would increase the listing performance more effectively for entire homes and shared rooms than for private rooms. For the private rooms, the slope was negative but not significantly different from zero. For the private rooms, the listing performance is improved with the increment of "interior style" and "attraction/transportation". For the shared rooms, the listing performance is better when the host mentions more about "kitchen" and "visitor". For all the listing types, the slope of "room" is also positive but weakly significant with the p-value slightly larger than 0.05 (0.0568). For the entire homes and the shared rooms, "attraction/transportation" and "interior style" have a negative impact on the listing performance. For the shared rooms, "building structure" is also negatively related to the listing performance.

In summary, for the entire homes, the descriptions about "kitchen", "building structure", and "room" have positive effects on the listing performance, while the descriptions about "interior style", "attraction/transportation", and "room" do the same for the private rooms. For the shared rooms, the description about "kitchen", "visitor" and "room" are positively related to the performance.

We find substantial differences in the most influential aspects as measured by the impact on listing performance and the most frequently mentioned aspects in listing descriptions. In Figure 3 and Table 5, we find that the hosts for entire homes mentioned "attraction/

transportation", "communication", and "building structure" the most often on average. Particularly, "kitchen" was mentioned the second least often though it has the most significant influence on listing performance. The hosts of private rooms were likely to mention "communication", "building structure", and "attraction/transportation" the most frequently, whereas "interior style" which was mentioned less frequently, was strongly and positively related to listing performance. For the shared rooms, the hosts wrote about "visitor", "room", and "communication" the most often on average. While "kitchen" was the most important aspect to increase the listing performance, we observed that it was not emphasized frequently by hosts in their descriptions.

4.3 Sensitivity analysis and model robustness

We conducted additional analyses to verify the robustness of our results. The estimation procedure was repeated with a linear mixed effect model using the same control and independent variables in Model 3. With cross-sectional data in each quarter, a linear model was also fitted. The parameter estimates were compared in Table 6. With significance level 0.1, all of the significant coefficients in Model 3 are also significant in the models fitted in Table 6, except for $D_{SharedRoom}$ and $D_{SharedRoom}{:}Comm$. In addition, the signs of the significant coefficients were consistent across the models compared, confirming the robustness of the results in Section 4.2.

Reverse causality and omitted variables are two potential sources of endogeneity in our models. We reduced the risk of reverse causality by using a lagged variable approach. Specifically, we modeled the panel data using the negative binomial mixed effect models where the independent variables in quarter *t-1* were used to model the dependent variable in quarter *t*.

In empirical studies, the omitted variable bias or misspecification of the model has been commonly checked by the coefficient movements when the regression model is modified by control variables (Chen and Pearl, 2015). A review of the literature in The American Economic Review (Lu and White, 2014) and other economic journals (Oster, 2013) finds that approaches such as sensitivity analysis of results with varying control variables are widely used for checking for omitted variable bias. Similarly, our sensitivity analysis results and the comparison of Models 1 and 2 have shown that the coefficient estimates do not vary significantly across the models, thus demonstrating the robustness of our models against omitted variable bias.

5. Discussion

In this section, we further elaborate on the findings and discuss their implications on practice and theory. In addressing our first research objective, we used an unsupervised deep learning-based aspect extraction technique and identified several common aspects in host-provided listing descriptions. Specifically, we extracted ten unique aspects commonly used by hosts in their listing descriptions. We find that in describing their listing, hosts focus on aspects related to the physical aspects of the property, its amenities, location-specific attributes, host-specific attributes and their expectations of the types of visitors to whom the offering is best suited.

There are noticeable differences in average aspect frequencies across the three different listing types, indicating that hosts customize listing descriptions to market segments. In the case of shared rooms, hosts emphasize the "room" aspect which describes the amenities available in the room being offered, and the "visitor" aspect which details the host expectations of the type of visitor for whom the shared room may be best suited. However, their emphasis on "attractions/transportation", "neighborhood" and "building structure" is

Table 6.
Sensitivity analysis

Fixed effect	Linear mixed effect model		Linear model with 2019.Q1		Linear model with 2019.Q2		Linear model with 2019.Q3		Linear model with 2019.Q4	
	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.
(Intercept)	5.884	(0.357)***	4.874	(0.540)***	5.810	(0.575)***	7.197	(0.601)***	5.613	((.480)***
D_privateroom	0.892	(0.421)*	1.563	(0.652)*	1.912	(0.690)**	1.800	(0.714)*	1.301	(0.560)*
D_sharedroom	1.078	(1.049)	1.909	(3.049)	1.061	(2.340)	2.179	(2.050)	0.234	(1.541)
Attr	−0.126	(0.074)	0.042	(0.108)	−0.040	(0.117)	−0.094	(0.121)	−0.005	(0.097)
Comm	0.050	(0.073)	0.123	(0.114)	0.078	(0.121)	0.035	(0.127)	0.154	(0.100)
Bldg	0.393	(0.080)***	0.406	(0.120)***	0.625	(0.130)***	0.628	(0.135)***	0.496	(0.111)***
Visitor	0.066	(0.093)	0.111	(0.135)	0.282	(0.153)	0.310	(0.161)	0.183	(0.127)
Room	0.023	(0.090)	0.112	(0.135)	0.154	(0.149)	0.043	(0.152)	0.001	(0.124)
Style	−0.262	(0.086)**	−0.240	(0.123)	−0.181	(0.136)	−0.281	(0.139)*	−0.242	(0.116)*
Nature	0.069	(0.088)	0.016	(0.129)	0.131	(0.143)	0.003	(0.145)	0.041	(0.122)
Nbhd	−0.032	(0.097)	−0.021	(0.140)	0.023	(0.154)	0.055	(0.157)	−0.006	(0.130)
Kitchen	0.569	(0.107)***	0.439	(0.153)**	0.719	(0.168)***	0.892	(0.176)***	0.759	(0.143)***
Facil	−0.002	(0.098)	0.004	(0.154)	−0.108	(0.160)	−0.180	(0.168)	−0.072	(0.134)
Downtown	0.773	(0.090)***	0.329	(0.120)**	0.583	(0.128)***	0.810	(0.134)***	0.758	(0.106)***
PerCapitaIncome	−0.365	(0.094)***	−0.341	(0.124)**	−0.471	(0.134)***	−0.582	(0.138)***	−0.229	(0.108)*
metrosf	0.256	(0.204)	0.747	(0.268)**	0.794	(0.284)**	0.520	(0.297)	0.217	(0.241)
review_scores_rating_avg	0.337	(0.067)***	−0.197	(0.129)	−0.034	(0.130)	0.154	(0.121)	0.091	(0.102)
Host_duration	−0.181	(0.037)***	−0.343	(0.054)***	−0.447	(0.056)***	−0.463	(0.057)***	−0.293	(0.044)***
$D_{Superhost}$	1.071	(0.105)***	2.655	(0.234)***	3.079	(0.254)***	3.112	(0.258)***	2.602	(0.212)***
$D_{PrivateRoom}$:*Attr*	0.541	(0.116)***	−0.068	(0.174)	0.497	(0.182)**	0.806	(0.190)***	0.590	(0.154)***
$D_{PrivateRoom}$:*Comm*	0.030	(0.109)	0.165	(0.174)	0.094	(0.182)	−0.069	(0.190)	0.025	(0.149)
$D_{PrivateRoom}$:*Bldg*	−0.270	(0.118)*	−0.307	(0.174)	−0.334	(0.189)	−0.523	(0.199)**	−0.323	(0.157)*
$D_{PrivateRoom}$:*Visitor*	−0.029	(0.136)	0.056	(0.205)	−0.063	(0.224)	−0.292	(0.239)	−0.154	(0.186)
$D_{PrivateRoom}$:*Room*	0.248	(0.148)	−0.381	(0.221)	−0.139	(0.236)	0.438	(0.250)	0.074	(0.196)
$D_{PrivateRoom}$:*Style*	0.506	(0.143)***	0.138	(0.205)	0.269	(0.223)	0.624	(0.231)**	0.603	(0.181)***
$D_{PrivateRoom}$:*Nature*	−0.088	(0.147)	−0.026	(0.223)	−0.264	(0.238)	−0.054	(0.247)	−0.210	(0.196)
$D_{PrivateRoom}$:*Nbhd*	−0.220	(0.156)	−0.296	(0.228)	−0.411	(0.246)	−0.169	(0.258)	−0.228	(0.207)
$D_{PrivateRoom}$:*Kitchen*	−0.695	(0.159)***	−0.243	(0.238)	−0.772	(0.256)**	−1.249	(0.270)***	−1.066	(0.215)***
$D_{PrivateRoom}$:*Facil*	−0.207	(0.156)	−0.171	(0.239)	−0.222	(0.248)	−0.302	(0.264)	−0.301	(0.206)
$D_{SharedRoom}$:*Attr*	−0.712	(0.434)	−1.377	(0.771)	0.165	(0.693)	−0.319	(0.744)	−1.312	(0.596)*

(continued)

Table 6.

Fixed effect	Linear mixed effect model Estimate	Linear mixed effect model Std. err.	Linear model with 2019.Q1 Estimate	Linear model with 2019.Q1 Std. err.	Linear model with 2019.Q2 Estimate	Linear model with 2019.Q2 Std. err.	Linear model with 2019.Q3 Estimate	Linear model with 2019.Q3 Std. err.	Linear model with 2019.Q4 Estimate	Linear model with 2019.Q4 Std. err.
$D_{SharedRoom}$:*Comm*	−0.460	(0.287)	−0.259	(0.568)	−0.404	(0.480)	−0.659	(0.498)	−0.081	(0.447)
$D_{SharedRoom}$:*Bldg*	−1.800	(0.480)***	−1.958	(0.842)*	−1.383	(0.784)	−3.019	(0.841)***	−1.328	(0.680)
$D_{SharedRoom}$:*Visitor*	0.572	(0.258)*	0.374	(0.657)	0.522	(0.504)	0.907	(0.537)	1.125	(0.415)**
$D_{SharedRoom}$:*Bedroom*	0.148	(0.346)	−0.210	(0.704)	−0.705	(0.712)	0.062	(0.660)	0.159	(0.505)
$D_{SharedRoom}$:*Style*	0.587	(0.648)	−0.218	(1.482)	−0.755	(1.148)	−0.932	(1.324)	−0.815	(1.071)
$D_{SharedRoom}$:*Nature*	0.955	(0.498)	0.170	(1.171)	−0.732	(1.070)	1.094	(0.935)	1.176	(0.849)
$D_{SharedRoom}$:*Nbhd*	0.040	(0.714)	1.875	(1.866)	1.189	(1.715)	−1.308	(1.850)	−0.615	(1.284)
$D_{SharedRoom}$:*Kitchen*	0.516	(0.521)	2.556	(1.079)*	2.666	(1.000)**	0.232	(1.616)	0.524	(0.907)
$D_{SharedRoom}$:*Facil*	0.068	(0.607)	−0.642	(1.302)	−0.871	(1.420)	0.744	(1.469)	−0.710	(1.037)

Note(s): ***: p-value < 0.001, **: 0.001 < p-value < 0.01, *: 0.01 < p-value < 0.05

much lower in shared home listings when compared to private rooms and entire homes. In the case of entire homes, we see that listing descriptions tend to emphasize “interior style”, “nature”, and “neighborhood” aspects. Entire homes are typically marketed to families and larger groups and therefore we see an increased focus on interior living areas and surrounding features compared to private rooms and shared rooms.

In exploring the impact of the various aspects on listing performance, we find that the impact of aspect mentions on listing performance is moderated by listing type. In the case of entire homes, “building structure”, “kitchen” and “room” have significant positive influence on listing performance. The most influential aspects for private rooms were “interior style”, “attraction/transportation”, and “room”, whereas “kitchen”, “visitor”, and “room” were the most influential aspects for shared rooms. Among the control variables, our results are consistent with previous research indicating that review score rating and superhost status are positively related to listing performance.

The findings above have multiple implications for practice. First, we find that there is a substantial mismatch between most frequently emphasized aspects by hosts and the most influential aspects that have an impact on listing performance. When writing their listing descriptions, hosts should focus on aspects for which the listing description is seen as a low cost and higher benefit information source. Moreover, the aspects emphasized need to be customized to the specific market segment. For example, hosts often emphasize nearby attractions and transportation facilities in their listing descriptions. However, it is negatively associated with listing performance for entire homes and shared rooms. While nearby attractions and transportation facilities are important aspects of a listing, entire home customers are more likely to be tourist families who often have access to other reliable sources for such information (for e.g. Google Maps) to inform their choices. However, the host description would often be the only and lowest cost information source for building structure, room amenities and kitchen-related aspects which we see are more influential in impacting listing performance for entire homes. The market for shared rooms as well typically consists of longer-term residents looking to share rooms with likeminded professionals and therefore may not necessarily be interested in nearby attractions but are rather looking for room and kitchen amenities which have a positive impact on listing performance.

In the case of private rooms, we observe that “interior style” has a significant positive influence on listing performance. This points toward the different market segment catered to by private rooms. Past research suggests that satiation with hotel accommodation service and decorations can push users to switch to peer to peer accommodations (Yan *et al.*, 2019). Private rooms are more likely to be considered as a substitute for traditional hotel spaces, and customers are therefore looking for aspects that provide information on privacy of the room, convenience of access, and interior styling and comforts to make a decision regarding their stay.

The findings of our study have significant implications for theory and the literature on prepurchase user information seeking and the impact of MGC on sales performance in the sharing economy. Whereas there are multiple studies that have evaluated post-purchase information such as consumer reviews and its impact on listing performance, our study evaluated listing descriptions to provide a prepurchase perspective into consumer decision-making. We see that aspects that have significant mentions in consumer reviews such as host communication, may not necessarily be influential when included in listing descriptions. Moreover, we see that the impact of aspects in listing descriptions can vary significantly based on the market segment. Given the length limitations of listing descriptions and limited cognitive resources available to consumers, we also see that when listing descriptions include redundant information or information on aspects not of interest to the consumer, it can have a negative impact on listing performance.

6. Conclusions

In this paper, we investigated the effects of host-provided listing descriptions on listing performance. First, we use text analytics to extract and identify key aspects highlighted by hosts in their listing descriptions. We then use negative binomial mixed effect models to analyze the impact of the different aspects on listing performance. We also evaluate the differential impact of the aspects on listing performance in three different market segments including entire homes, private rooms and shared rooms. The context of our study is the Airbnb marketplace in San Francisco and Oakland. We use data from InsideAirbnb.com for listing related information including host-provided descriptions and customer ratings, and additional socioeconomic and geographic variables from Census and GIS datasets.

The results of our study have multiple implications for theory and practice. First, our findings confirm that host product descriptions have a significant impact on listing performance. This finding has significant implications in the sharing economy and indicates that hosts can use product descriptions as a valid marketing tool to impact sales. Second, we find that influential aspects vary across market segments and the findings, consistent with economics of information theory, indicate that users rely on host-provided listing descriptions for those aspects where hosts are uniquely best suited to provide the information and for which other sources of information are not available. This finding contributes to enhance our theoretical understanding of prepurchase consumer information search behavior. Third, we find that there is a mismatch in the aspects frequently emphasized by hosts and the most influential aspects. Finally, the approach we present in this paper which involved deep learning-based aspect extraction and regression analysis serves as an example that can be also used in other domains to identify and extract influential aspects across customer segments.

There are some limitations to our study. The research context of our study is limited to Airbnb listings in two cities, namely San Francisco and Oakland. Therefore, some aspects of our finding may not be generalizable to listing in remote or rural areas where the listings and customer segments may be different. Even among urban areas, although we have analyzed many listings, our analysis is limited to two cities and may not capture variation in cities in terms of density, economic activity and tourist activity. However, our approach can be easily replicated in different settings. In addition, the sample size of shared rooms is far less than those of entire homes and private rooms. This could have caused the listing descriptions of shared rooms to be less represented in the extracted aspects, and the uncertainty of the regression coefficient estimates was greater for shared rooms. Therefore, we limited our interpretations for shared rooms when the data points for aspects were limited.

Future research opportunities in this area include the independent application of text-mining techniques to different listing types for in-depth analysis of each individual market segment. In this paper, we extracted aspects of listing descriptions provide by hosts across the listing types. However, the results of our study and recent research indicate variations in influential aspects and customer expectations (Lee, 2021; Xu, 2020) based on listing type. In future research, we plan to modify the ABAE-based approach to extract aspects for each listing type independently to derive deeper insight into each individual listing type.

In summary, our contributions in this paper are two-fold. First, we provide practical guidance to hosts on influential aspects in different market segments. The findings in this study on influential aspects by market segments can benefit hosts in better targeting their customer segments and improving their listing performance. Second, we contribute to the literature on the impact of MGC on listing performance. Our findings inform future research and theories on the impact of MGC on listing performance, including the types of aspect that are likely to be influential and the moderating influence of market segments.

References

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F. and Iglesias, C.A. (2017), "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", *Expert Systems with Applications*, Vol. 77, pp. 236-246.

Bird, S., Loper, E. and Klein, E. (2009), *Natural Language Processing with Python*, O'Reilly Media.

Biswas, B., Sengupta, P. and Chatterjee, D. (2020), "Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques", *Decision Support Systems*, Elsevier B.V., Vol. 135, 113324.

Blei, D.M. and McAuliffe, J.D. (2007), "Supervised topic models", *Advances in Neural Information Processing Systems 20 – Proceedings of the 2007 Conference*, Vancouver, British Columbia.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, doi 10.1016/b978-0-12-411519-4.00006-9.

Chattopadhyay, M. and Mitra, S.K. (2019), "Do Airbnb host listing attributes influence room pricing homogenously?", *International Journal of Hospitality Management*, Vol. 81, doi 10.1016/j.ijhm.2019.03.008.

Chen, B. and Pearl, J. (2015), "Exogeneity and robustness", available at: http://ftp.cs.ucla.edu/pub/stat_ser/r449.pdf.

Chen, T., Xu, R., He, Y. and Wang, X. (2017), "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN", *Expert Systems with Applications*, Vol. 72, pp. 221-230.

Cheng, M. and Jin, X. (2019), "What do Airbnb users care about? An analysis of online review comments", *International Journal of Hospitality Management*, Vol. 76, pp. 58-70.

Dann, D., Teubner, T. and Weinhardt, C. (2019), "Poster child and Guinea pig – insights from a structured literature review on Airbnb", *International Journal of Contemporary Hospitality Management*, Vol. 31 No. 1, pp. 427-473.

Dimoka, A., Hong, Y. and Pavlou, P.A. (2012), "On product uncertainty in online markets: theory and evidence", *MIS Quarterly*, Vol. 36 No. 2, p. 395.

Eckhardt, G.M., Houston, M.B., Jiang, B., Lamberton, C., Rindfleisch, A. and Zervas, G. (2019), "Marketing in the sharing economy", *Journal of Marketing*, Vol. 83 No. 5, pp. 5-27.

Ert, E. and Fleischer, A. (2019), "The evolution of trust in Airbnb: a case of home rental", *Annals of Tourism Research*, Vol. 75, doi 10.1016/j.annals.2019.01.004.

Gunter, U. (2018), "What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay area", *Tourism Management*, Vol. 66, doi 10.1016/j.tourman.2017.11.003.

Gursoy, D. (2019), "A critical review of determinants of information search behavior and utilization of online reviews in decision making process (invited paper for 'luminaries' special issue of International Journal of Hospitality Management)", *International Journal of Hospitality Management*, Elsevier, Vol. 76, pp. 53-60.

Gursoy, D. and McCleary, K.W. (2004), "Travelers' prior knowledge and its impact on their information search behavior", *Journal of Hospitality and Tourism Research*, SAGE Publications, Vol. 28 No. 1, pp. 66-94.

He, R., Lee, W.S., Ng, H.T. and Dahlmeier, D. (2017), "An unsupervised neural attention model for aspect extraction", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Stroudsburg, Pennsylvania, Vol. 1, pp. 388-397.

Hothorn, T., Bretz, F. and Westfall, P. (2008), "Simultaneous inference in general parametric models", *Biometrical Journal*, Vol. 50 No. 3, pp. 346-363.

Jakob, N. and Gurevych, I. (2010), "Extracting opinion targets in a single- and cross-domain setting with conditional random fields", *EMNLP 2010 – Conference on Empirical Methods in*

Natural Language Processing, Proceedings of the Conference, Cambridge, Massachusetts, pp. 1035-1045.

Jin, W. and Ho, H.H. (2009), "A novel lexicalized HMM-based learning framework for web opinion mining", *Proceedings of the 26th Annual International Conference on Machine Learning – ICML'09*, ACM Press, New York, New York, pp. 1-8.

Lawani, A., Reed, M.R., Mark, T. and Zheng, Y. (2019), "Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston", *Regional Science and Urban Economics*, Vol. 75, pp. 22-34.

Lee, C.K.H. (2021), "How guest-host interactions affect consumer experiences in the sharing economy: new evidence from a configurational analysis based on consumer reviews", *Decision Support Systems*, 113634, doi: 10.1016/j.dss.2021.113634.

Lee, C.K.H., Tse, Y.K., Zhang, M. and Ma, J. (2019), "Analysing online reviews to investigate customer behaviour in the sharing economy", *Information Technology and People*, Vol. 33 No. 3, pp. 945-961.

Li, X., Wu, C. and Mai, F. (2019), "The effect of online reviews on product sales: a joint sentiment-topic analysis", *Information and Management*, Vol. 56 No. 2, pp. 172-184.

Liang, S., Schuckert, M., Law, R. and Chen, C.C. (2017), "Be a 'superhost': the importance of badge systems for peer-to-peer rental accommodations", *Tourism Management*, Vol. 60, doi 10.1016/j.tourman.2017.01.007.

Liang, S., Schuckert, M., Law, R. and Chen, C.C. (2020), "The importance of marketer-generated content to peer-to-peer property rental platforms: evidence from Airbnb", *International Journal of Hospitality Management*, Elsevier, Vol. 84, 102329.

Liu, A.X., Steenkamp, J.B.E.M. and Zhang, J. (2018), "Agglomeration as a driver of the volume of electronic word of mouth in the restaurant industry", *Journal of Marketing Research*, Vol. 55 No. 4, pp. 507-523.

Lu, X. and White, H. (2014), "Robustness checks and robustness tests in applied economics", *Journal of Econometrics*, Vol. 178 No. Part 1, doi 10.1016/j.jeconom.2013.08.016.

Luo, L., Ao, X., Song, Y., Li, J., Yang, X., He, Q. and Yu, D. (2019a), "Unsupervised neural aspect extraction with sememes", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization*, California, pp. 5123-5129.

Luo, Y., Tang, R. and Liang (2019b), "Understanding hidden dimensions in textual reviews on Airbnb: an application of modified latent aspect rating analysis (LARA)", *International Journal of Hospitality Management*, Vol. 80, pp. 144-154.

Lutz, C. and Newlands, G. (2018), "Consumer segmentation within the sharing economy: the case of Airbnb", *Journal of Business Research*, Vol. 88, pp. 187-196.

Manyika, J., Lund, S., Bughin, J., Robinson, K., Mischke, J. and Mahajan, D. (2016), "Independent work: choice, necessity, and the gig economy", available at: https://www.mckinsey.com/featured-insights/employment-and-growth/independent-work-choice-necessity-and-the-gig-economy.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013), "Distributed representations of words and phrases and their compositionality", *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada.

Mimno, D. and McCallum, A. (2008), "Topic models conditioned on arbitrary features with Dirichlet-multinomial regression", *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI 2008*, Helsinki.

Mimno, D., Wallach, H.M., Talley, E., Leenders, M. and McCallum, A. (2011), "Optimizing semantic coherence in topic models", *EMNLP 2011 – Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, Edinburgh, pp. 262-272.

Minka, T. and Lafferty, J. (2002), "Expectation-propagation for the generative aspect model", *UAI'02: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, Edmonton, pp. 352-359.

Nieto García, M., Muñoz-Gallego, P.A., Viglia, G. and González-Benito, Ó. (2019), "Be social! The impact of self-presentation on peer-to-peer accommodation revenue", *Journal of Travel Research*, Vol. 59 No. 7, pp. 1268-1281, doi 10.1177/0047287519878520.

Nisar, T.M., Hajli, N., Prabhakar, G. and Dwivedi, Y. (2019), "Sharing economy and the lodging websites", *Information Technology and People*, Vol. 33 No. 3, pp. 873-896.

Oster, E. (2013), "Unobservable selection and coefficient stability: theory and validation", NBER Working Paper No. 19054, National Bureau of Economic Research.

Roberts, M.E., Stewart, B.M. and Airoldi, E.M. (2016), "A model of text for experimentation in the social sciences", *Journal of the American Statistical Association*, Taylor and Francis, Vol. 111 No. 515, pp. 988-1003.

Röder, M., Both, A. and Hinneburg, A. (2015), "Exploring the space of topic coherence measures", *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining – WSDM'15*, ACM Press, New York, New York, pp. 399-408.

Rosario, A.B., Sotgiu, F., De Valck, K. and Bijmolt, T.H.A. (2016), "The effect of electronic word of mouth on sales: a meta-analytic review of platform, product, and metric factors", *Journal of Marketing Research*, Vol. 53 No. 3, doi 10.1509/jmr.14.0380.

Smith, A. (2016), "Shared, collaborative and on demand: the new digital economy", available at: https://www.pewresearch.org/internet/2016/05/19/the-new-digital-economy/.

Stigler, G.J. (1961), "The economics of information", *Journal of Political Economy*, Vol. 69 No. 3, doi 10.1086/258464.

Tussyadiah, I.P. and Park, S. (2018), "When guests trust hosts for their words: host description and trust in sharing economy", *Tourism Management*, Vol. 67, pp. 261-272.

Wang, L., Liu, K., Cao, Z., Zhao, J. and de Melo, G. (2015), "Sentiment-aspect extraction based on restricted Boltzmann machines", *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp. 616-625.

Wang, W., Pan, S.J., Dahlmeier, D. and Xiao, X. (2016), "Recursive neural conditional random fields for aspect-based sentiment analysis", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp. 616-626.

Xie, K. and Mao, Z. (2017), "The impacts of quality and quantity attributes of Airbnb hosts on listing performance", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 9, pp. 2240-2260.

Xinhua (2020), "China's sharing economy to grow by up to 15 pct annually: report", 7 March, available at: http://www.xinhuanet.com/english/2020-03/07/c_138852084.htm.

Xu, X. (2020), "How do consumers in the sharing economy value sharing? Evidence from online reviews", *Decision Support Systems*, North-Holland, Vol. 128, 113162.

Yan, R., Zhang, K.Z.K. and Yu, Y. (2019), "Switching from hotels to peer-to-peer accommodation: an empirical study", *Information Technology and People*, Vol. 32 No. 6, pp. 1657-1678.

Ye, Q., Law, R., Gu, B. and Chen, W. (2011), "The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings", *Computers in Human Behavior*, Vol. 27 No. 2, pp. 634-639.

Zhang, J. (2019a), "Listening to the consumer: exploring review topics on Airbnb and their impact on listing performance", *Journal of Marketing Theory and Practice*, Vol. 27 No. 4, pp. 371-389.

Zhang, J. (2019b), "What's yours is mine: exploring customer voice on Airbnb using text-mining approaches", *Journal of Consumer Marketing*, Vol. 36 No. 5, pp. 655-665.

Zhang, L., Yan, Q. and Zhang, L. (2018), “A computational framework for understanding antecedents of guests’ perceived trust towards hosts on Airbnb”, *Decision Support Systems*, North-Holland, Vol. 115, pp. 105-116.

Zhang, M., Zhang, Y. and Vo, D.T. (2015), “Neural networks for open domain targeted sentiment”, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp. 612-621.

Corresponding author
Surendra Sarnikar can be contacted at: surendra.sarnikar@csueastbay.edu