Decision Support Systems 135 (2020) 113324

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques

Baidyanath Biswas[a,*], Pooja Sengupta[b], Dwaipayan Chatterjee[c]

[a] Information Technology Area, International Management Institute, Kolkata 700027, India
[b] Quantitative Methods Area, International Management Institute, Kolkata 700027, India
[c] Department of Statistics, University of Calcutta, Kolkata 700019, India

ARTICLE INFO

Keywords:
Sharing economy
Airbnb
Poisson
Quasi-Poisson
Negative Binomial
Text analytics

ABSTRACT

The sharing economy has experienced massive growth in the short-term shared-home rental industry. However, few studies have investigated the determinants of the number of customer reviews received by these shared-homes. To fill this gap, we were motivated to propose an analytical framework that identified these determinants, both explicit and implicit. We applied Poisson, Quasi-Poisson, and Negative Binomial regressions with a dataset consisting of Airbnb properties from ten different cities worldwide, while successful bookings were proxied by the count of customer reviews posted by guests. We performed a cluster analysis based on the properties to generate homogeneous "cluster cities" and performed the regressions separately for each cluster. Among host-generated features, *superhost*, *host duration*, *bedrooms*, and *amenities* became significant. Among user-generated features, *overall review scores* and *negative sentiments* were significant. We also found that the "superhost" badge moderated the effects of host-generated content on the count of customer reviews. Consequently, guests paid a higher "price per night" for "superhost" properties, while they overlooked crucial attributes such as "website features." Through these novel "cluster-specific" recommendations, our study extends the existing theories and contributes to the literature of decision analytics and tourism management. Finally, we performed a sensitivity analysis to check for the timeliness and robustness of these determinants.

1. Introduction

The *sharing economy* also referred to as *collaborative consumption*, is a peer-to-peer (P2P) mode of subscribing, sharing, or attaining access to products and services that is often enabled by a community-based digital platform [1–3]. In recent years, the sharing economy has rapidly grown in the tourism and hospitality sectors, especially for short-term shared-home rentals [4,5]. Online platforms such as Airbnb, VRBO, and HomeAway have pioneered the use of shared accommodation facilities by connecting people who own unused assets (such as unoccupied rooms, apartments, and houses) with those who require short-term rentals (such as tourists and business visitors), via digital marketplaces [6–8]. Although these two-sided platforms closely resemble traditional electronic marketplaces, shared-homes aim to create additional value [3,9] by offering short-term rentals at cheaper rates[1], and with much more extensive assortments of accommodation assets than traditional hotels and lodges [10,11]. Therefore, P2P home-sharing platforms are aiming to offer a new value proposition [12] through *a large variety of selection*, *cheaper rates*, and *mutual trust*. However, the actual booking data is not publicly available, and so scholars have adopted the "count of customer reviews" as an alternative measure [3,13,14]. Accordingly, our study focuses on the number of reviews received by shared-homes, as this is the primary indication that the properties were indeed rented to guests and reflect end-to-end P2P transactions.

Every property on the P2P home-rental platform has its distinct characteristics, and every host has a distinct personality. For Airbnb, its potential consumers need to exert relatively more efforts to choose from a diversity of homes that may range from private rooms to castles[2],

* Corresponding author.
E-mail address: b.biswas@imi-k.edu.in (B. Biswas).

[1] Is Airbnb Really Cheaper Than A Hotel Room in The World's Major Cities? Retrieved From: https://www.forbes.com/sites/niallmccarthy/2018/01/23/is-airbnb-really-cheaper-than-a-hotel-room-in-the-worlds-major-cities-infographic/#664a8c3078ac

[2] From Treehouses to Aluminum Pods - 10 of the Most Unique Homes on Airbnb, Retrieved From: https://news.airbnb.com/en-ca/from-treehouses-to-aluminum-pods-10-of-the-most-unique-homes-on-airbnb/

https://doi.org/10.1016/j.dss.2020.113324
Received 1 November 2019; Received in revised form 18 May 2020; Accepted 18 May 2020
Available online 22 May 2020
0167-9236/

whereas the inventory is standardized in the case of hotels. Previously, scholars have examined the effects of user-generated content (UGC) (such as *product reviews*) on the sellers' online sales in e-commerce [15,16]. Further, the UGCs posted by consumers are based on their experiences and are more likely to be perceived as trustworthy. However, for P2P rental platforms such as Airbnb, the marketer-generated-content (MGC) is no less important to consumers [14]. These MGCs include objective information about the online platform (such as *self-description of property*), property features (such as *exact location*, *bedrooms*, *amenities*)[3], the *price per night*, rental policies (such as *instant bookable*, *cancellation rules*, *phone verification of guest*), availability, and finally the location. Fig. 1 presents a P2P shared-home as shown on Airbnb. Often, users interpret some of these features *explicitly*, while others may remain *implicit*. For instance, the *title of a property*, as displayed on the P2P website, is an *explicit* MGC, whereas its *readability* and *expressivity* are *implicit* MGCs. Similarly, historical reviews are an *explicit* UGC, while their *sentiment polarities* and *cognitive insights* are *implicit* UGC features. Further, the social exchange theory explains social behaviour exhibited during the interaction among two parties [17]. A cost-benefit analysis follows that can be measured in the form of *customer reviews* and their antecedents - MGC and UGC features, emotions, and perceived evaluations [3]. Thus, we pose our first research question: **RQ1- What are the major (UGC and MGC based) determinants of the count of customer reviews on Airbnb?**

Often the nature of urban tourism varies with each city such that P2P shared-homes have different occupancy rates. The expectation of arriving tourists vary significantly, and so does the behaviour of hosts. Therefore, each city has a unique pattern of the count of customer reviews received by the Airbnb properties, as well as their underlying explanatory factors[4]. Wang and Nicolau [18] reported heterogeneity effects of the pricing determinants of Airbnb properties across different cities. We propose a clustering mechanism based on the number of P2P properties across homogeneous groups, such that each "cluster of cities" remains consistent. Additionally, Airbnb can offer unique cluster-based recommendations to their hosts and attract more customer reviews instead of forcing a *one-size-fits-all* strategy. Thus, we pose our second research question: **RQ2- How do these determinants influence the count of customer reviews on Airbnb across different clusters?**

Again, when two parties trade in P2P marketplaces without any face-to-face interactions, the hosts can experience severe damage to their properties while the guests may also encounter the risk of untrustworthy hosts, thereby leading to economic and privacy risks. In contrast, traditional hotels and lodges display relatively higher levels of trust because they can reduce the above risks through standardization, safety regulations, and business reputation [19]. After a fatal accident at an Airbnb property in 2019, Brian Chesky, the CEO of Airbnb, has said, "ultimately, we're in the business of trust."[5] Therefore, P2P shared-homes require robust trust-building mechanisms and two-way safety verifications to encourage trustworthiness among the guests as well as hosts that can help to mitigate the associated risks[6]. Efficient reputation mechanisms installed in Airbnb include the *superhost* feature [20], *profile photos* [21,22], *verified host identity* [23], and *membership duration* on the platform [3,24]. Meanwhile, the self-theory suggests that individuals regularly adjust their self impressions to attain respective goals in daily lives [25]. Hence, consumers perceive these reputation systems and self-descriptions [23,26] as a "trust factor" that moderates the MGC features of an Airbnb property during successful sales (measured by the count of customer reviews). So far, extant studies have focused on isolated settings, simulated user environments, and often without actual review data from Airbnb. While a few scholars have identified the factors of perceived trust, in this study, we aim to investigate the effect of trust on the host-generated features that may influence the customer reviews. Thus, we pose our third research question: **RQ3 - How does trust moderate the MGCs of the count of customer reviews on Airbnb?**

To address these research questions, we propose an analytical framework to identify the MGC and UGC determinants of the number of customer reviews received by successful shared-properties, understand their relative significance across clusters, and finally examine the moderating effect of trust-based factors that can influence those reviews. First, we extracted the MGC details of each property and combined them with the reviews (UGC) posted. Accordingly, in this study, we intended to measure the successful sales by the number of customer reviews posted by guests after their stays at Airbnb properties[7]. Second, we applied advanced text-mining to generate implicit determinants from linguistic cues. This technique enabled us to avoid multicollinearity issues that could arise while using similar independent variables as well as incorporate customer behaviours while posting reviews and build our explanatory models. Third, we categorized the Airbnb property data according to homogeneity in the "number of listings," leading to four unique clusters. This clustering mechanism enabled us to explore the effects of the review determinants for each cluster separately and propose unique sets of recommendations for the Airbnb hosts. Fourth, we applied count data regressions, namely, Poisson, Quasi-Poisson, and Negative Binomial, to develop our analytical framework. Fifth, we ran a series of variable importance schemes and fine-tuned them to include only the top ten variables for each cluster. Finally, to perform a sensitivity analysis on the timeliness and representativeness of the explanatory variables, we re-ran our regression models within a reduced timeframe starting from 2014 and onwards.

The remainder of this paper is organized as follows. In Section 2, we present an overview of the existing academic works and theoretical foundations on peer-to-peer home rentals in the sharing economy and explore their determinants. In Section 3, we present the methodology, describe, explore the data, and introduce our proposed analytical framework. In Section 4, we present the results obtained from the explanatory analytical framework and perform an additional round of sensitivity analysis. In Section 5, we discuss the significance of the results, followed by their academic and managerial implications in Section 6. Finally, in Section 7, we conclude this study and highlight the future scope for an extension.

2. Literature review and theoretical foundation

2.1. Peer-to-peer home rentals in the sharing economy

P2P home-sharing applications are two-sided digital platforms that resemble electronic marketplaces and facilitate social benefits through the principles of platform economics [4]. Dann et al. [27] had conducted an extensive review of the literature on Airbnb articles that were published between 2013 and 2018. Their study inferred some major themes that included, among others, customer motives, trust and reputation systems, online reviews, and pricing determinants. We have examined the extant literature in this domain, and we have identified two broad perspectives: i) *customers* and ii) *hosts*, that closely resemble the classification into (i) *marketer* or *host*-generated content (MGC), and (ii) *user*-generated content (UGC), as proposed by Liang et al. [14].

[3] https://www.airbnb.com/hospitality

[4] https://www.statista.com/statistics/752498/airbnb-number-of-listings-in-major-cities-worldwide/

[5] https://www.cnbc.com/2019/11/06/airbnb-is-pushing-major-safety-changes.html

[6] https://blog.atairbnb.com/building-trust/

[7] https://www.airbnb.com/help/article/13/how-do-reviews-work-for-stays

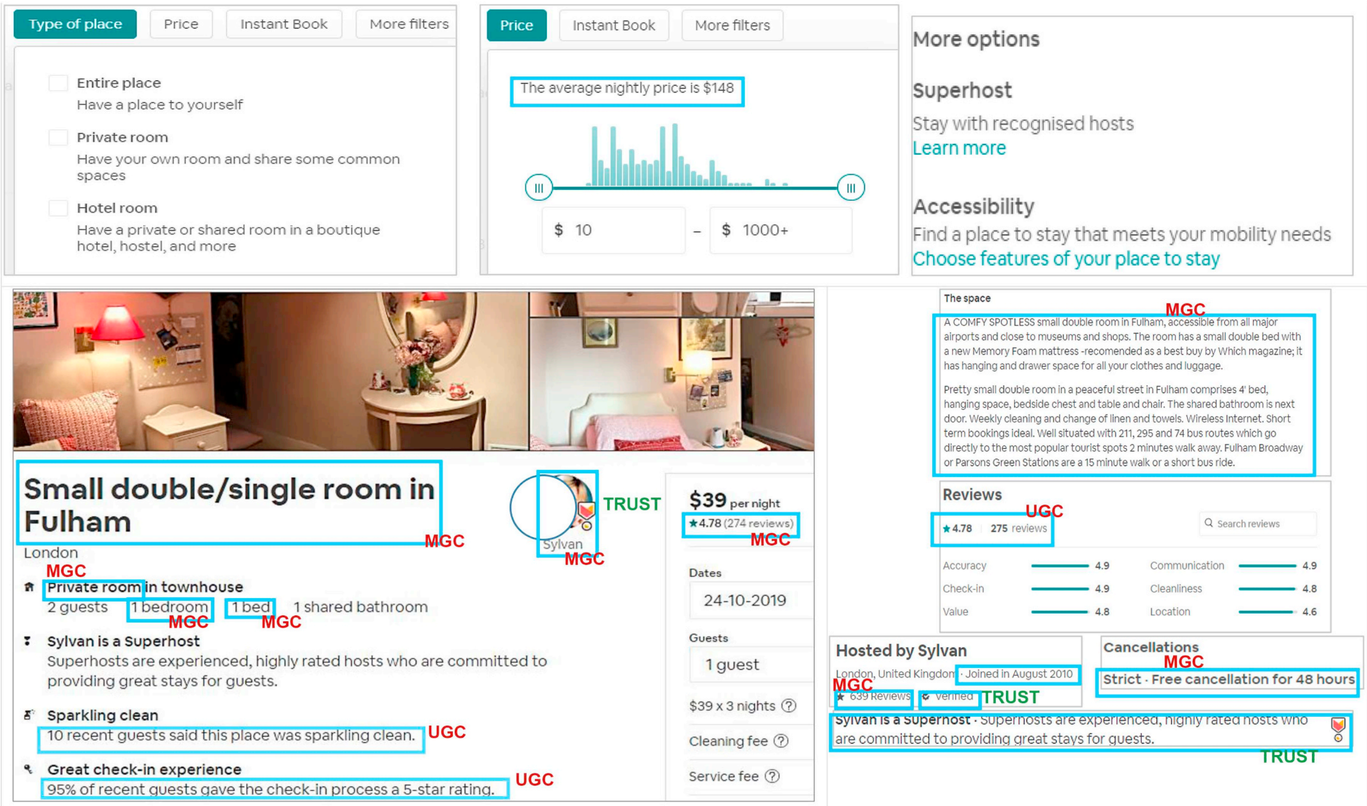


Fig. 1. A typical shared-home rental on Airbnb platform with UGC and MGC features

2.2. *Determinants of customer reviews received by peer-to-peer shared-homes*

We categorized the determinants of the number of customer reviews received by successful P2P transactions on the Airbnb platform into MGC and UGC. These were further subdivided into (i) website (MGC), (ii) host (MGC), (iii) property (MGC), (iv) historical reviews (UGC), (v) rental policies (MGC), and (vi) availability (MGC).

Website Attributes: The Airbnb platform behaves as the primary interface between the host and the guest. Therefore, the title and self-description of a property [14], readability [13], and expressivity of description as measured through the use of superlatives [28,29], are significant MGC-based determinants of the count of customer reviews received by an Airbnb property. Bilgihan and Bujisic [30] noted that good design features of the hotel website are essential for quicker room reservations, followed by a higher volume of reviews. Mauri et al. [31] found that the presence of storytelling narratives within host profiles could increase the popularity of the property by at least 8% and increase the number of reviews received. Recently, Liang et al. [14] reported the effects of website-based host-generated content (or MGC) to improve the volume of reviews for an Airbnb property. Therefore, we posit that hosts can increase the count of customer reviews by providing legible and comprehensive descriptions related to their shared-homes.

Host Attributes: In the traditional hospitality industry, hotel ownerships and chain affiliations (such as Mariott, Hyatt) can signal the quality of a property [32]. However, in the case of Airbnb, there is no standardized measure of the quality of the property and host reputation. Instead, Airbnb has adopted the following two indicators – (i) *host identity verification* through connections within their social networks, govt. IDs, self-disclosures [26], and personal photos [21,22], (ii) *certification* through the "Superhost" badge granted by Airbnb. According to Airbnb, a host attains the "Superhost" badge: (i) if a host has completed a minimum of 10 stays annually, (ii) quickly responds to the queries, (iii) has at least 4.8 average ratings, and (iv) has less than 1% cancellation rate [8]. Recent studies have investigated the effects of the "Superhost" badge, "host identity verification," and reported their positive outcomes on the volume of customer reviews received by Airbnb properties [20,33–35]. Referring to the previous studies, we included the duration as a determinant of the count of customer reviews, where it represents the length of time for which a host has been active on Airbnb, and measured by the number of months [3,24]. In this way, these host-based antecedents or MGC-s can improve host-branding [23] and also signal trust mechanisms to the customer [9].

Property Attributes and Pricing: Among the property attributes of MGC, we identified the *exact location, bed type, number of bedrooms, room type, amenities offered*, and *price per night*. Ert et al. [21] noted that the apartment-size and the location of Airbnb homes had a significant impact on their purchase decisions. Recent studies have also found that the location of P2P shared-homes can impact their subsequent sales [24,28]. Liang et al. [20] found that shared-homes with fewer bedrooms but more beds, receive a higher volume of reviews. Qiu et al. [36] and Liang et al. [14] also noted that when an entire property was available or had real beds, its chances of being booked could increase manifold, leading to a higher count of customer reviews. Guttentag [10], Chattopadhyay and Mitra [35], and Lyu et al. [37] noted that Airbnb listings with amenities such as wireless Internet, free parking, kitchen, and laundry services received a higher volume of reviews as compared to those properties that lacked them. A mass of academic literature has examined the determinants of pricing using the property-based attributes offered in Airbnb homes [35]; [18]. Kwok and Xie [38] noted the effects of Airbnb pricing strategies on the performance of hotels in the neighbourhood. Masiero et al. [39] found that the location

[8] https://www.airbnb.com/Superhost

[9] Perfect Strangers: How Airbnb is building trust between hosts and guests, Retrieved From: https://news.airbnb.com/perfect-strangers-how-airbnb-is-building-trust-between-hosts-and-guests/

could positively influence hotel-room prices, whereas the distance from the city centre had a negative effect on the hotel-room bookings and the ensuing reviews. Guttentag and Smith [10] noted that the Airbnb rooms showed a significant displacement effect in comparison to the adjacent hotels while the tourists selected rooms. Accordingly, we find that the property attributes and pricing can largely determine the count of customer reviews received by Airbnb properties.

Historical Review Attributes: With the advent of UGC, the guest (or the customer) has begun to post a summary narrating the overall perception of the property. Extant literature in the hospitality and tourism industry has focused on online reviews and customer feedback [40–42]. Recently, Chen and Chang [43] and Ju et al. [44] had identified the purchase intention of potential customers on Airbnb through an examination of review ratings. Chen and Jin [45] had extracted the positive and negative keywords from the review-text and subsequently studied their influences on the guest's overall experience. In addition to the sentiments, we propose that the prospective customers shall consider linguistic cues [29] to decide upon their future bookings, that is measured by the count of customer reviews. These linguistic cues are embedded in the review-text and constitute important implicit determinants, such as cognitive emotional contents that are measured by certainty (indicated by words like *think, know*) and insights (indicated by words like *always, never*).

Rental Policies and Availability: Among the MGC attributes based on the rental policies, we chose *instant bookable*, *strict* versus *other cancellation policies*, and whether the property *required the guest's phone number* to approve the reservation. Liang et al. [14] and Wang and Nicolau [18] noted that the rental and cancellation policies of shared-homes[10] played a significant role in price-setting, room-booking, and subsequently, the customer reviews received by the Airbnb property. While some guests preferred to search and book at the last minute, others searched and booked far in advance. Airbnb also recommends to its hosts that it is a good idea to plan at least 2-3 months in advance[11]. Hence, we consider the *90-day property availability* as a significant determinant of customer reviews received by Airbnb properties.

2.3. Theoretical background

The theoretical background of this study lies primarily in the social penetration theory [46], the social exchange theory [17], and the value co-creation theory [12]. Social penetration theory explicates the evolution of interpersonal relationships, which become stronger and more reliable when individuals disclose themselves to one another over time [46]. It also validates the significance of self-disclosure and mutual exchanges in developing and consolidating a relationship [47]. Thus, self-disclosure and mutual interactions help alleviate the information asymmetry among the guests and hosts, thereby enhancing customer satisfaction and future consumption intentions [21,22,26]. Social penetration theory also supports that the reciprocal P2P interactions between the hosts and guests are the reasons for generating the electronic word-of-mouth intentions, experiencing satisfaction, and behaviour to re-use shared-homes [48].

Social exchange theory explains the emotional and social behaviours that ensue an interaction between two parties. For a social exchange to be successful, the "parties must abide by specific 'rules' of exchange," noted by Emerson [17] as the "normative definition of the situation that forms among, or is adopted by the participants in an exchange relationship." Social exchange theory also suggests that a cost-benefit analysis occurs during the host-guest interactions [3,49]. While the hosts invest their accumulated "reputational" capital (such as the "Superhost" badge) in setting the price per night [14,20,23,33], the guests expect enjoyment (*amenities, bedrooms*, and *room type*) [36,37], economic benefits (*lowered prices*) [18,35], and satisfaction (manifested as *positive reviews, review certainty*, and *insights*) [14,43,49,50]. Price negotiation acts as a controlling mechanism throughout the entire social-cum-economic exchange.

Scholars suggest that a service-provider cannot merely offer value as embedded in goods and services, but gradually "co-create" value through an intrinsically collaborative exercise [12,51]. The service-dominant (S-D) logic introduced initially by Vargo and Lusch [12], has become the pioneering model in value co-creation studies. In line with S-D logic, it follows that instead of considering value as pre-existing, "the value can only be created with and determined by the user in the 'consumption' process and through use" [12]. Linking P2P property-rentals with the S-D framework, we note that Airbnb offers a unique value proposition as it enables direct encounters with the hosts as well as provides access to the physical resources, all in the presence of human actors to form a basis for the resultant value co-creation [10,49,52,53]. Subsequently, the co-created value can be encouraged through a "value proposition" when using intangible competencies (such as consumer reviews, trust, and reputation) as well as tangible elements (such as host's house, amenities, rental policies, and website features) to "make them more valuable" [12]. To summarize, the social penetration of interpersonal relationships occur when guests stay in P2P shared-homes, followed by an exchange of perceptions among the hosts and guests, and a cost-benefit analysis when they experience a value co-creation during their interactions. These theories serve as the foundation for our study and help us in developing the determinants for our framework. Fig. 2 illustrates our proposed conceptual framework with the key determinants of the count of customer reviews received by a P2P shared-home.

3. Methodology

3.1. Data description and feature engineering

We collected the data for the P2P shared-homes and their reviews posted by customers from insideairbnb.com,[12] a third-party website that collects data from Airbnb. Based on a recent update [13] on insideairbnb.com, we examined shared-properties within the period ranging from September 2010 until January 2020. Fig. 3 illustrates the geo-location of each listed property from London, as shown on insideairbnb.com. We considered both UGC and MGC determinants to examine their effects on the dependent variable (i.e., the posting of reviews after successful Airbnb stays). Among those determinants, some were explicit (both *numerical* and *categorical*), while others were implicit. These implicit variables were based on linguistic cues, such as (i) *readability of the self-description, superlatives used* (MGC); (ii) *positive sentiments, negative sentiments, review certainty, review insights* (UGC) [44], and were generated using the LIWC (Linguistic Inquiry and Word Count) software [54]. For instance, a long but illegible description may discourage readers, unless both the length and the readability are suitable. In this study, we computed the readability as $Readability = 1/GunningFogIndex$, where $(GunningFogIndex) = 0.4 * (ASL + PHW)$, ASL = the average sentence length, and PHW = the percentage of compound words [13]. Next, we counted the number of adjectives used to describe the property on the Airbnb website, such as *a perfect room*, or *airy space*, which would invariably attract a prospective consumer. We calculated it with the help of Stanford Natural Language Toolkit (NLTK). Positive words (such as *love, beautiful, sweet*) and negative words (such as *hurt, ugly, nasty*) attributed to the particular sentiments and were calculated using LIWC. Cognitive dimensions embedded in

[10] https://www.airbnb.com/home/cancellation_policies

[11] https://blog.atairbnb.com/experiences-pricing-availability-tips/

[12] http://insideairbnb.com/get-the-data.html

[13] Most cities listed on insideairbnb.com receive minor file updates every month, our study considered the property-based data up to January 2020

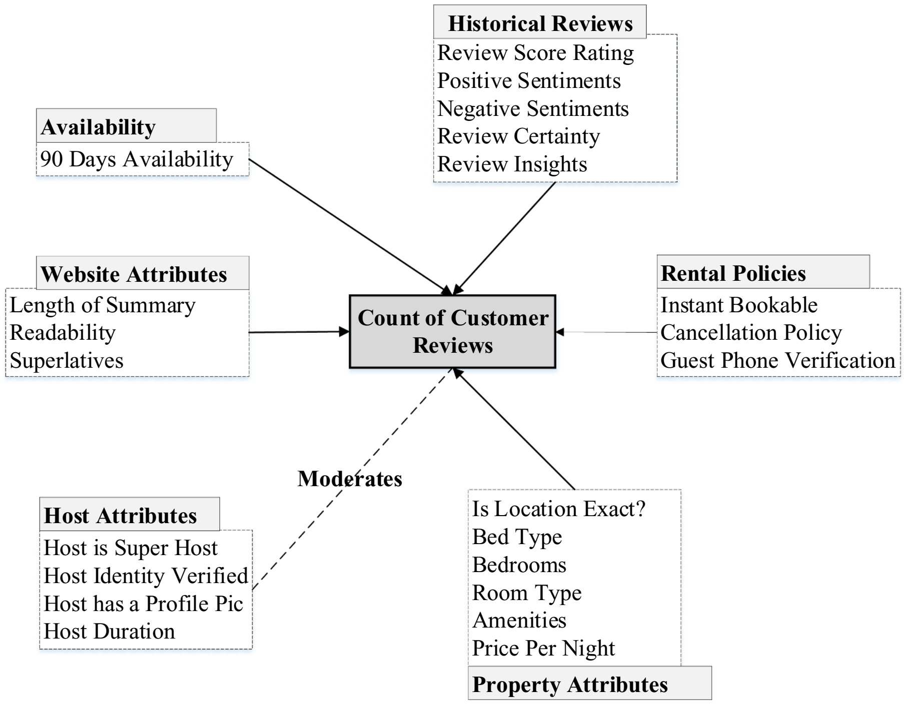


Fig. 2. Proposed conceptual framework.

the reviews - (i) insights (such as *think, know*), and (ii) certainty (such as *always*, *never*) were also calculated using LIWC.

Finally, we listed the independent variables from each category: *website, host, property, reviews, rental policies*, and *availability*, similar to Walls et al. [55], described them in brief, and presented their descriptive statistics in Table 1. Our categorization also matches with findings from the text-regression results reported by Xu [3]. In our study, we have used the "count of customer reviews" posted by the guests after their successful stays as a proxy of bookings, because (i) Airbnb reservation data for each property is not available publicly, (ii) some scholars have already adopted the "count of customer reviews" as an alternative measure of actual reservations [3,13,14]. From Table 1,

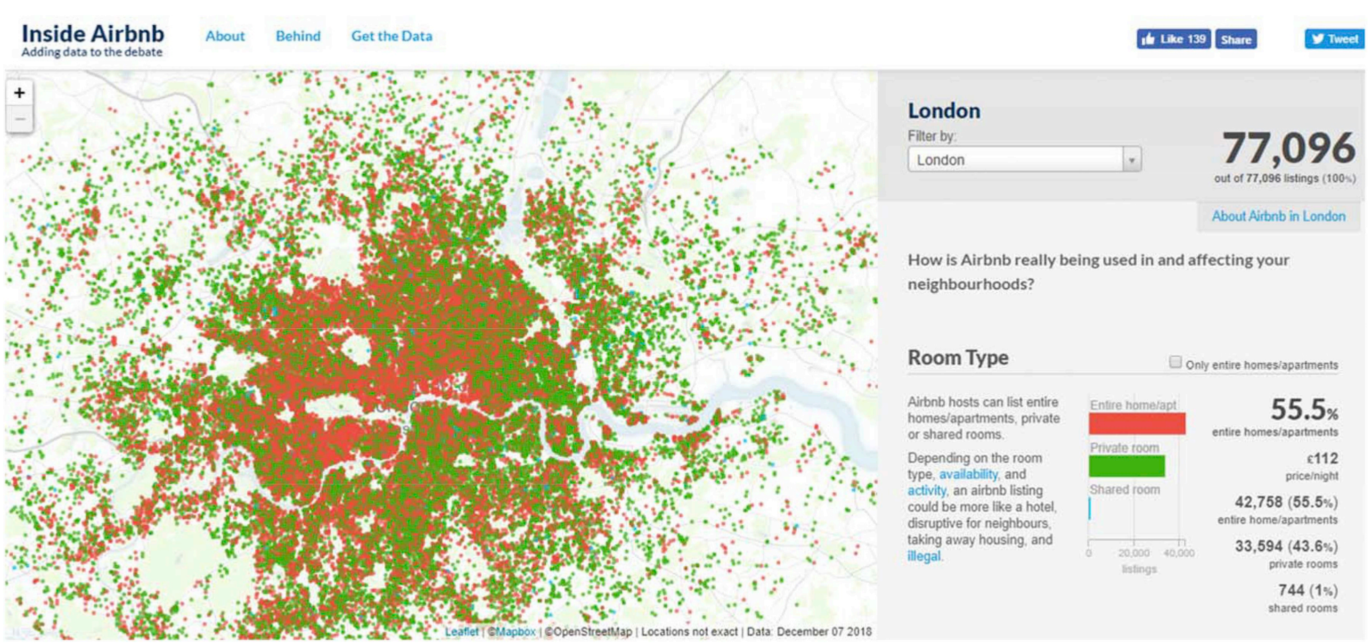


Fig. 3. The geo-location of each listing in London.
Red – entire home/apartment, Green – private room. 77,096 properties at the end of 2018 with 55.5 % shown as entire home/apartment (Source: insideairbnb.com).

Table 1
Brief description of variables used in our analytical framework.

	Variable	Type	Definition	Literature Source	Theory	Mean	SD	Max	Min
Independent									
	Length of Summary	M	Length (in words) of property summary (N)	Developed from [13,14]	SET	54.91	27.48	196	0
Website	Readability	M	Text readability of property description (N)	[13,14]	SET	0.08	0.02	0.43	0.05
	Superlatives	M	Adjectives used in the title of property description (N)	Self Developed	SET	8.12	11.62	22	0
	Host is Superhost	M	Whether host is a super host (D)	[14,20]	SPT	-	-	-	-
Host	Host Identity Verified	M	Whether identity of host is verified (D)	[26]	SPT	-	-	-	-
	Host has a Profile Pic	M	Whether host's profile picture is displayed (D)	[21,22]	SPT	-	-	-	-
	Host Duration	M	How long (in months) host is present on Airbnb (N)	[3,24]	SPT	47.40	31.04	113.90	0
	Location Exact	M	Whether given location of property is accurate or not (D)	Self Developed	SET	-	-	-	-
	Bed Type	M	Offers a real bed (versus airbed, sofa, futon) (D)	[14,36]	SET	-	-	-	-
Property	Bedrooms	M	Number of bedrooms available in property (N)	[34]	SPT	1.55	1.12	46	0
	Room Type	M	Offers an entire home (versus hotel, prviate, shared) (D)	[34]	SPT	-	-	-	-
	Amenities	M	Number of amenities present at property (N)	[3,24]	SET	23.30	11.50	109	0
	Price per Night	M	Per night price of property (US$) (N)	[18,35]	SET	174.50	453.05	36127	0
	Review Scores Rating	U	Overall avg. review scores received (N)	[34,43]	SET, VCT	93.50	9.56	100	0
Historical	Positive Sentiments	U	Overall avg. positive polarity of reviews received (N)	Self Developed	SET, VCT	11.51	5.85	100	0
Reviews	Negative Sentiments	U	Overall avg. negative polarity of reviews received (N)	Self Developed	SET, VCT	1.67	0.94	17.76	0
	Certainty	U	Certainty in cognition of reviews received (N)	Self Developed	SET, VCT	2.19	2.01	100	0
	Insights	U	Insights from reviews received (N)	Self Developed	SET, VCT	0.72	0.81	50	0
	Instant Bookable	M	Whether instant booking is available (D)	[36]	SET, VCT	-	-	-	-
Rental	Cancellation Policy	M	Whether cancellation policy is strict (or otherwise) (D)	Developed from [18]	SET, VCT	-	-	-	-
Policy	Guest Phone Verification	M	Whether property requires guest phone to approve (D)	Self Developed	SET, VCT	-	-	-	-
Availability	Availability 90	M	Availability of property for next 90 days (N)	Developed from [24]	SET, VCT	36.36	34.8	90	0
Dependent									
	Customer Reviews	U	Number of reviews received by property (N)	[3,14]	SET, VCT	32.367	110.464	879	1

N = Numeric; D = Dummy; U = Based on User-Generated-Content; M = Based on Marketer-Generated-Content.
SET = Social Exchange Theory; VCT = Value Co-Creation Theory; SPT = Social Penetration Theory.

we observed that some determinants exhibited varying scales and much higher ranges than others, such as the *host duration* and the *number of reviews*. Further, some variables suffered from high standard deviations, such as the *length of the summary*, the *price per night*, and *bedrooms*. Therefore, to improve the empirical results and ensure the accuracy of the coefficient estimations, we decided to normalize the research variables [3,20] and log-transform them before model-fitting [14]. Fig. 4 illustrates these methodological steps that were adopted in our study.

3.2. Clustering

We chose Airbnb listings from ten different cities across the world, namely, Brussels, Mallorca, San Diego, Austin, Melbourne, Toronto, Montreal, Los Angeles, Sydney, and London. These cities have a varying number of property listings on the Airbnb platform. Due to these differences in listings, we posit that there are changing patterns in the number of customer reviews as well as the relevant factors that affect them. For instance, Wang and Nicolau [18] have examined the pricing determinants of Airbnb homes for a sample of 180,533 properties chosen from 33 cities, which were aggregated together to fit a single empirical model. In contrast, Xie and Mao [56] have empirically estimated the effects of host-attributes on the listing performance (which was measured by the number of successful reservations in subsequent months) and were moderated by the number of listings for the city of Austin, Texas. Hence, a *one-size-fits-all* strategy cannot be adopted because that might yield highly absurd and non-actionable results for the Airbnb hosts residing in these cities. Therefore, we performed a cluster analysis to form homogeneous groups of cities based on their respective number of listings, instead of including them as a control variable

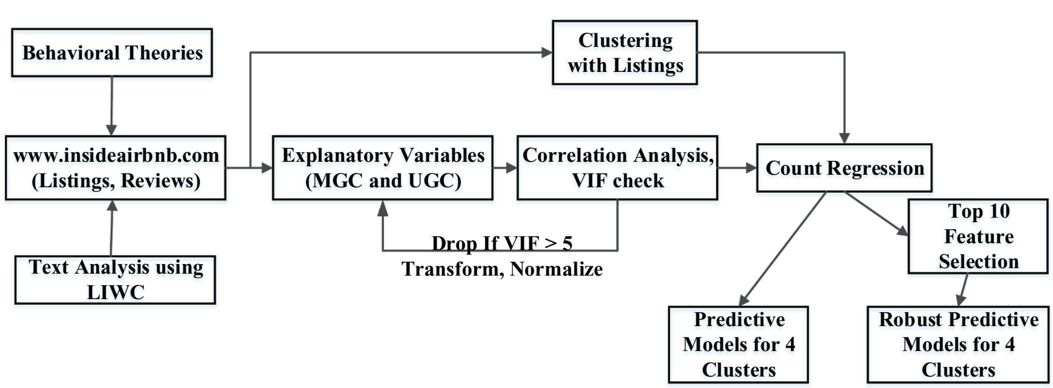


Fig. 4. Methodological steps to execute our analytical framework

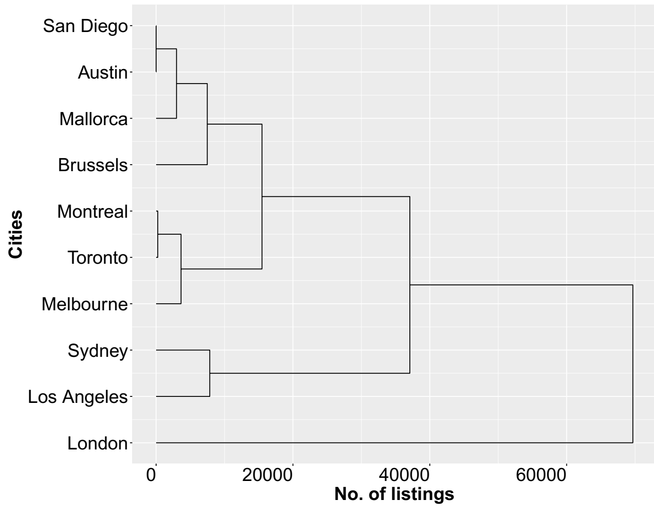


Fig. 5. A dendrogram for the cluster analysis of cities based on the number of Airbnb properties

during the empirical analysis. This step ensured that each cluster remained homogeneous in terms of the number of listings. Then we fitted the empirical models for each cluster separately to find the determinants of the count of customer reviews among them. Intending to collect useful insights from the Airbnb count data, we used these clusters to extract the natural homogeneity of each "cluster-city." Figure 5 illustrates the dendrogram with four clusters that we built from the aggregate Airbnb listing data.

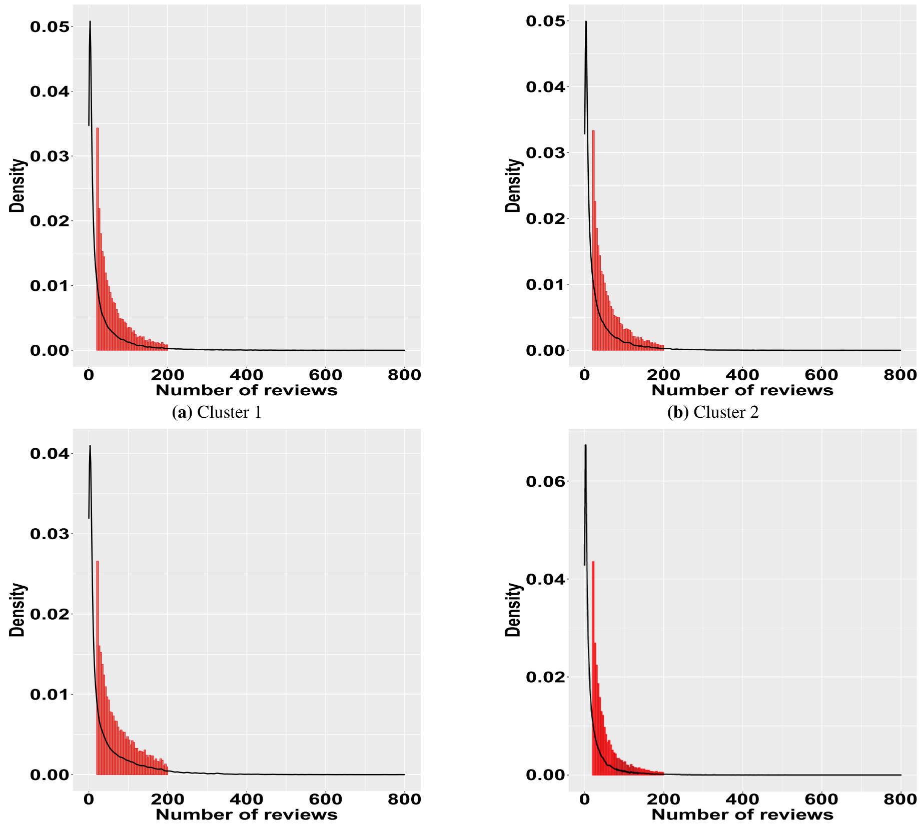
The proposed clustering generated four groups: *Cluster 1* was composed of Austin, Brussels, Mallorca, and San Diego ($N = 49{,}822$), which had the least number of listings. *Cluster 2* was composed of Melbourne, Toronto, and Montreal ($N = 68{,}380$); *Cluster 3* was composed of Los Angeles and Sydney ($N = 78{,}462$); *Cluster 4* was composed of London ($N = 85{,}068$) where the density of Airbnb listings was the highest across all cities in the world. In this way, the unique cluster analysis applied in this study helped us to empirically analyze the count of customer reviews received by Airbnb properties and perform an in-depth investigation of their MGC and UGC attributes from different cities.

3.3. *Pairwise correlation among variables and multicollinearity checking*

Before proceeding with the model-fitting with Airbnb property data, we computed the pairwise correlations among the 23 variables selected for our proposed framework. Then we checked whether the variance inflation factor (VIF) stayed within the permissible limit of 10. For Cluster 1, the pairwise correlation values ranged between 0.349 and -0.338, while the VIF values ranged from 1.000 to 1.346. For Cluster 2, the pairwise correlation values ranged between 0.403 and -0.344, while the VIF values ranged from 1.001 to 1.353. For Cluster 3, the pairwise correlation values ranged between 0.395 and -0.320, while the VIF values ranged from 1.000 to 1.742. For Cluster 4, the pairwise correlation values ranged between 0.330 and -0.211, while the VIF values ranged from 1.000 to 1.255. For each cluster, we found that the pairwise correlation values were much lower than the permissible limit of 0.5, and none of the variables had VIF values above 5. Thus, the pairwise correlations and VIFs for the variables used in this study were well within the allowable ranges for each cluster. As we intend to build separate explanatory models for each cluster 1, 2, 3, and 4 using the 22 variables, we checked the pairwise correlations and VIF-s separately for each cluster. However, due to space limitations, we reported the results for London (Cluster 4) only, in Table 2. Fig. 4 presents the associated steps for executing the VIF and correlation checks.

Table 2
Pairwise correlation among variables from Cluster 4 (85,068 observations).

	VIF	[01]	[02]	[03]	[04]	[05]	[06]	[07]	[08]	[09]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]
Summary Length [01]	1.051	1																						
Readability [02]	1.000	0.001	1																					
Superlatives [03]	1.001	-0.019	0.002	1																				
Superhost [04]	1.148	0.065	-0.008	0.004	1																			
Host Id Verify [05]	1.145	0.017	-0.006	0.023	0.089	1																		
Host Profile Pic [06]	1.000	0.002	0.002	-0.011	0.003	0.003	1																	
Host Duration [07]	1.216	-0.025	0.001	0.027	0.067	0.330	0.006	1																
Location Exact [08]	1.010	0.027	0.003	0.004	0.042	0.029	-0.006	0.052	1															
Bed Type [09]	1.006	0.015	0.001	-0.004	-0.015	-0.034	0.002	-0.046	-0.011	1														
Bedrooms [10]	1.171	0.056	0.005	0.001	-0.032	0.013	0.004	0.034	0.015	0.045	1													
Room Type [11]	1.218	0.025	0.145	0.040	-0.008	0.045	-0.004	0.109	-0.043	-0.007	0.037	1												
Amenities [12]	1.210	0.188	-0.005	-0.001	0.254	0.078	0.001	0.046	-0.010	0.005	0.179	0.008	1											
Price per Night [13]	1.120	0.026	0.005	-0.002	-0.020	-0.041	0.002	-0.042	-0.025	0.025	0.299	0.089	0.085	1										
Review Scores Rating [14]	1.140	0.032	0.004	0.001	0.179	0.068	-0.004	0.077	0.038	-0.008	0.038	0.017	0.090	0.004	1									
Positive Sentiments [15]	1.255	0.015	-0.001	0.002	0.068	0.004	0.008	-0.003	0.010	0.002	-0.047	-0.140	0.028	-0.015	0.241	1								
Negative Sentiments [16]	1.120	-0.007	-0.002	0.001	-0.001	0.004	0.001	0.001	0.003	0.001	-0.006	0.231	-0.009	-0.001	0.013	0.319	1							
Certainty [17]	1.081	0.005	-0.006	0.003	0.045	0.006	-0.007	0.008	0.007	-0.004	-0.010	0.089	0.027	0.012	0.150	0.253	0.063	1						
Insights [18]	1.007	-0.007	-0.006	-0.001	0.026	0.012	0.006	0.018	0.008	-0.008	-0.019	-0.211	0.006	-0.013	0.062	-0.007	-0.016	0.043	1					
Instant Book [19]	1.078	0.027	-0.007	-0.019	-0.034	-0.145	-0.001	-0.208	-0.057	0.040	-0.048	-0.011	0.029	0.028	-0.100	-0.006	-0.003	-0.013	-0.024	1				
Cancellation Policy [20]	1.080	0.072	0.007	-0.008	0.020	0.039	0.002	0.029	-0.026	0.017	0.133	0.105	0.186	0.098	-0.040	-0.017	-0.006	0.003	-0.009	0.046	1			
Guest Phone Verify [21]	1.056	-0.010	0.004	0.005	0.036	0.072	-0.004	0.194	0.001	-0.011	0.009	0.004	0.086	0.002	-0.006	-0.008	-0.002	0.001	0.013	0.001	0.077	1		
Availability 90 [22]	1.078	-0.049	-0.003	-0.006	0.045	-0.049	-0.001	-0.061	-0.033	-0.003	-0.016	-0.021	0.127	0.098	-0.080	0.003	-0.007	-0.008	-0.003	0.076	0.143	0.048	1	
Customer Reviews [23]	1.097	0.009	0.006	0.010	0.212	0.080	-0.004	0.106	0.029	-0.030	-0.094	-0.155	0.130	-0.052	0.052	0.037	-0.006	0.021	0.016	0.030	0.054	0.086	0.105	1



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4

Fig. 6. Density plots for "number of reviews".

3.4. Empirical modelling

Modelling count data is a challenging analytical task in economics and social sciences. Scholars adopt Poisson regression as the benchmark technique to model count data and often prefer it to the common practice of log transformation of the dependent variable [58]. In Poisson regression, we assume that the Poisson incidence rate of μ is determined by a set of k regressors (i.e., the *X's*). Therefore, if we want to model the count of customer reviews received by an Airbnb property with the help of those regressors (from Table 1), let y_i be the count of customer reviews for the i^{th} property and y_i follows a Poisson distribution with mean μ_i such that: response $y_i \sim \mathrm{Po}(\mu_i)$. Hence the linear component in the regression model is used to parametrize the mean rate of customer reviews μ_i, and is given by the functional form:

$$\eta_i = \log \mu_i = \beta_{0i} + \beta_{i1} x_{i1} + \cdots + \beta_{ik} x_{ik} \tag{1}$$

We use the Poisson regression model to fit the expected number of customer reviews for the i^{th} property, as follows:

$$\mathrm{E}[y_i] = \mu_i = \exp(\eta_i) \tag{2}$$

where the link function $g(\mu_i) = \log(\mu_i)$ is the (natural) logarithm. In Eqn. (1), β_{i0} is the intercept, and the regression coefficients $\beta_{i1}, \beta_{i2}, \beta_{i3}, \ldots, \beta_{ik}$ are unknown parameters for the i^{th} cluster of cities, which we intend to estimate from the available Airbnb data. The final estimates of these coefficients are labelled as $b_{i1}, b_{i2}, b_{i3}, \ldots, b_{ik}$.

The assumption that the mean and variance are equal (as in Poisson regression) may not hold for a particular count dataset. This anomaly is commonly known as the *overdispersion effect* [58]. To remediate this problem, scholars have often used a scaled Poisson regression, which works as a quasi-likelihood method. This Quasi-Poisson model is capable of handling overdispersed count data from Airbnb in our study. From Figure 6, we observed that the density plots for the dependent variable *"number of reviews"* were non-normal [20]. We further validated the presence of overdispersion by comparing the mean and variance as well as through the Dean's Overdispersion Test [57], which we reported in Table 3. We assumed i.i.d. frequency of the customer reviews at Airbnb: $N_1, N_2, \ldots N_m$, where μ_i is the mean such that $\mu_i = E(N_i)$. We then defined a log-link function $g(\mu_i) = \mathit{log}\,(\mu_i)$ (similar to a regular Poisson) to build the Quasi-Poisson model. Again, quasi-likelihood models require a specific relationship between mean and variance instead of the detailed theoretical assumptions on the distributions [58]. Our Quasi-Poisson model adopted its mean-variance relationship from the Poisson distribution such that the variance was related to the mean by a multiplicative factor of dispersion ϕ, given as: $Var(N_i) = \phi V(\mu_i) = \phi\mu_i$; where N_i is the count of customer reviews and $V(\mu_i)$ is the variance function in the generalized linear model (GLM) setting.

Next, we applied Negative Binomial regression to solve for overdispersion [58]. The equation for Negative Binomial (NB) regression is

Table 3
Dean's Overdispersion Test for "number of reviews".

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Mean	32.778	32.037	42.765	22.804
Variance	3253.009	2625.690	4681.547	1642.334
Dean's Statistic† [57]	10645***	11222***	16419***	11386***

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; $\dagger H_{alt}: Var > Mean$

similar to Cameron and Trivedi [59], such that $y_i \sim \mathrm{NB}(\mu_i)$, and is of the form:

$$\log \mu_i = \eta_i = \beta_{i0} + \beta_{i1}x_{i1} + \cdots + \beta_{ik}x_{ik} \tag{3}$$

Explanatory Model for RQ1 and RQ2: Therefore, the regression equation used in our study to explain the count of customer reviews received by Airbnb properties across clusters is of the form:

$$\begin{gathered}\eta_i\,[\mathit{Count\ of\ Customer\ Reviews}] = \beta_0 + \beta_1 \log(\text{Length of Summary}) \\ + \beta_2\,\text{Readability} + \beta_3\,\text{Superlatives} + \beta_4\,\text{Superhost} + \beta_5\,\text{Identity Verified} \\ + \beta_6\,\text{Profile } \mathit{Pic} + \beta_7\,\text{Duration} + \beta_8\,\text{Location Exact} \\ + \beta_9\,\mathit{Bed}\ \text{Type} + \beta_{10}\,\mathit{log}(\text{Bedrooms}) + \beta_{11}\,\text{Room Type} \\ + \beta_{12}\,\text{Amenities} + \beta_{13}\,\mathit{log}(\text{Price } \mathit{per} \text{ night}) + \beta_{14}\,\text{Review Scores} \\ + \beta_{15}\,\text{Positive Sentiment} + \beta_{16}\,\text{Negative Sentiment} + \beta_{17}\,\text{Review Certainty} \\ + \beta_{18}\,\text{Review Insights} + \beta_{19}\,\text{Instant Bookable} + \beta_{20}\,\text{Cancellation Policy} \\ + \beta_{21}\,\text{Guest Phone Verified} + \beta_{22}\,\text{Availability 90}\end{gathered}$$

Explanatory Model for RQ3: To address the moderating effect of "trust," we considered interactions between the reputation system (i.e., "superhost") and MGC-s (such as "website" and "property"), while the effects of UGC-s on the customer reviews remained unchanged. Therefore, the regression equation for the count of customer reviews is of the form:

$$\begin{gathered}\eta_i\,[\mathit{Count\ of\ Customer\ Reviews}] = \beta_0 + \beta_1\,\text{Readability} + \beta_2\,\text{Superlatives} \\ + \beta_3\,\text{Location Exact} + \beta_4\,\mathit{Bed}\ \text{Type} \\ + \beta_5\,\mathit{log}(\text{Bedrooms}) + \beta_6\,\text{Room Type} + \beta_7\,\text{Amenities} \\ + \beta_8 \log(\text{Price per Night}) + \beta_9\,\text{Superhost} \\ + \beta_{10}\,\text{Superhost} * \text{Superlatives} + \beta_{11}\,\text{Superhost} * \text{Readability} \\ + \beta_{12}\,\text{Superhost} * \text{Location Exact} + \beta_{13}\,\text{Superhost} * \text{Bed Type} \\ + \beta_{14}\,\text{Superhost} * \log(\text{Bedrooms}) + \beta_{15}\,\text{Superhost} * \text{Room Type} \\ + \beta_{16}\,\text{Superhost} * \text{Amenities} + \beta_{17}\,\text{Superhost} * \log(\text{Price per Night}) \\ + \beta_{18}\,\text{Positive Sentiments} + \beta_{19}\,\text{Negative Sentiments} \\ + \beta_{20}\,\text{Certainty} + \beta_{21}\,\text{Insights}\end{gathered}$$

3.5. Feature selection and variable importance scoring

In data-intensive analytical problems dealing with statistical computing, variable importance schemes are executed after fitting the appropriate regression techniques. Often these schemes lead to a better understanding of the independent variables as well as their effects on the dependent variable while building the regression models [35]. In this study, we proposed an analytical framework (Fig. 2) that uses 22 independent variables (Table 1) to predict the count of customer reviews. We applied the *varImp()* function for GLM setting from the *caret* package in R to refine our prediction models and retain the top 10 important features. For parametric models, the *varImp()* evaluation function uses model information to obtain the importance of the variables. In general, an LM/GLM model-based approach is closely dependent on the model performance. Sometimes it can even incorporate the correlation structure between the predictors to calculate the variable importance. In this way, *varImp()* creates an aggregate importance score by applying the absolute value of the t-statistic for each model parameter and then ranks the features accordingly. In our study, the Poisson, Quasi-Poisson, and Negative Binomial models were fitted using GLM, and thus we obtained the variable importance scores that helped in subsequent feature selection.

4. Empirical results

We conducted our experiments in the open-source R environment on a Windows system with an Intel core and 1.8 GHz processor. We used a blend of regression models to estimate the main effects and robustness checks. For each cluster, we applied the Poisson, followed by Quasi-Poisson, and Negative Binomial regression models with the *glm()* function. We present the explanatory models to examine the count of customer reviews received by an Airbnb property using the Poisson, Quasi-Poisson, and Negative Binomial regression in Tables 4, 5, 6, and 7 for Clusters 1, 2, 3 and 4 respectively. For each cluster, Poisson and Negative Binomial regression had similar parameter estimations but large differences in the "goodness of fit" due to overdispersion effects. Therefore, we applied the Quasi-Poisson and Negative Binomial regression models for each cluster separately to handle overdispersion.

Results show that the coefficients of *length of summary*, *readability*, *superlatives*, *superhost status*, *host identity verified*, the *host has a profile picture*, *host duration* on the Airbnb platform, *exact location* of the property, *amenities offered*, *review scores ratings*, *positive sentiments of reviews received*, *cognitive certainty of review*, *cognitive insights of reviews received*, *instant bookable*, *guest phone verification*, and *90-day availability* are consistently positive, significant and comparable for the main results across all the four clusters. Whereas, the coefficient estimates of *bed type*, *number of bedrooms*, *room type*, the *price per night*, *negative sentiments of reviews received*, and *cancellation policy* are consistently negative and, therefore, inversely affect the count of customer reviews received by a P2P shared-home on the Airbnb platform.

We then computed the importance scores for the features that influence the number of customer reviews. Based on our proposed methodological steps (Fig. 4), first, we built the count regression models using all the 22 independent features and then with the top 10 features only, which were derived from the importance scheme. We performed this exercise for each cluster using Negative Binomial regression and reported them next to their original coefficient estimates in Tables 4, 5, 6, and 7. Fig. 7 presents these scores for the top 10 features across each cluster. We observe that there is a significant change in the top 10 ranked-set across clusters. Results confirm that our proposed analytical framework is valid and efficient for each cluster while using the entire set of 22 features (as listed in Table 1) as well as for the top 10 features (as shown in Fig. 7).

Next, we built the explanatory models to answer for RQ3, where we addressed the moderating effect of "trust" through interactions between the Airbnb reputation system (i.e., "superhost") and MGC-s (such as "website" and "property"), while the effects of the UGC-s on the customer reviews remained unchanged. We present these results in Table 8 such that *NB-Ci* represents the Negative Binomial model fitted with data from Cluster *Ci*, where i = 1,2,3,4. Coefficient estimates and significance values of the variables to examine the count of customer reviews received by an Airbnb property as well the interaction terms, confirmed that trust in the form of "superhost" badge acts as a moderator to enhance the effects of a few variables only. For example, the *readability* and *superlatives* had lower coefficient estimates and were of lesser significance while the interaction terms for *bed type*, *number of bedrooms*, *room-type*, and *price per night* had higher coefficient estimates. However, the coefficient estimates for the *exact location* and *amenities* did not show any significant difference at all.

In an experiment, sensitivity analysis allows the researcher to identify whether any likely bias has highly influenced the results and if the findings are truly robust. Because of problems that could arise from the long time-span of the explanatory variables in our analytical framework, we performed a sensitivity analysis on their timeliness and representativeness. Therefore, we re-organized the explanatory variables on an annual basis and built models under the framework of longitudinal data analysis while considering Airbnb properties from $t = 2014$ onwards. Table 9 presents the results with a revised set of observations for each cluster given as follows, Cluster 1: $N = 36{,}305$ instead of $N = 49{,}822$; Cluster 2: $N = 52{,}364$ instead of $N = 68{,}380$; Cluster 3: $N = 59{,}168$ instead of $N = 78{,}462$; and Cluster 4: $N = 64{,}114$ instead of $N = 85{,}068$. We observed that the coefficient estimates and significance for the *readability*, *superhost*, *host duration*, and *negative sentiments* had increased over time, while the estimates and significance for the *room type* and *host identity verified* had decreased over time. The estimates remained unaffected across clusters for some

Table 4
Explanatory models to examine the count of customer reviews from Airbnb Cluster 1.

	Dependent variable: Count of customer reviews			
	Poisson	Quasi-Poisson	Neg Bin All	Neg Bin Top10
log (Length of Summary)	$0.001^{**}(0.0003)$	$0.001^{**}(0.0003)$	$0.0005^{***}(0.0002)$	
Readability	$0.004^{***}(0.0002)$	$0.004^{***}(0.0002)$	$0.003^{***}(0.0001)$	
Superlatives	$0.004^{***}(0.0001)$	$0.004^{***}(0.0001)$	$0.004^{***}(0.0001)$	
Superhost	$0.409^{***}(0.0020)$	$0.409^{***}(0.0180)$	$0.343^{***}(0.0140)$	$0.488^{***}(0.0130)$
Host Identity Verified	$0.159^{***}(0.0020)$	$0.159^{***}(0.0050)$	$0.053^{***}(0.0110)$	$0.050^{**}(0.0120)$
Host has a Profile Pic	$0.250^{***}(0.0570)$	$0.271^{***}(0.0490)$	$0.281^{***}(0.0107)$	$0.277^{**}(0.1050)$
Host Duration	$0.003^{***}(0.0040)$	$0.003^{***}(0.0040)$	$0.006^{***}(0.0026)$	$0.006^{***}(0.0001)$
Location Exact	$0.236^{***}(0.0220)$	$0.236^{***}(0.0170)$	$0.168^{***}(0.0180)$	$0.208^{***}(0.0001)$
Bed Type	$-0.333^{***}(0.0070)$	$-0.310^{***}(0.0059)$	$-0.351^{***}(0.0057)$	
log(Bedrooms)	$-0.536^{***}(0.0120)$	$-0.530^{***}(0.0118)$	$-0.598^{***}(0.0094)$	$-0.655^{***}(0.0013)$
Room Type	$-0.118^{***}(0.0020)$	$-0.118^{***}(0.0019)$	$-0.236^{***}(0.0140)$	
Amenities	$0.017^{***}(0.0018)$	$0.016^{***}(0.0022)$	$0.015^{***}(0.0045)$	$0.014^{***}(0.0004)$
log(Price per Night)	$-0.202^{***}(0.0020)$	$-0.202^{***}(0.0065)$	$-0.133^{***}(0.0070)$	$-0.180^{***}(0.0060)$
Review Scores Rating	$0.010^{***}(0.0001)$	$0.010^{***}(0.0005)$	$0.024^{***}(0.0001)$	
Positive Sentiments	$0.005^{***}(0.0002)$	$0.005^{***}(0.0010)$	$0.010^{***}(0.0001)$	
Negative Sentiments	$-0.008^{***}(0.0015)$	$-0.007^{***}(0.0010)$	$-0.012^{***}(0.0001)$	
Certainty	$0.020^{***}(0.0003)$	$0.020^{***}(0.0030)$	$0.055^{***}(0.0030)$	
Insights	$0.044^{***}(0.0010)$	$0.044^{***}(0.0009)$	$0.087^{***}(0.0007)$	
Instant Bookable	$0.175^{***}(0.0020)$	$0.175^{***}(0.0014)$	$0.142^{***}(0.0050)$	$0.144^{**}(0.0304)$
Cancellation Policy	$-0.375^{***}(0.0050)$	$-0.362^{***}(0.0041)$	$-0.289^{***}(0.0020)$	
Guest Phone Verification	$0.166^{***}(0.0040)$	$0.149^{***}(0.0360)$	$0.154^{***}(0.0340)$	
Availability 90	$0.004^{***}(0.0002)$	$0.004^{***}(0.0002)$	$0.003^{***}(0.0001)$	$0.005^{***}(0.0001)$
Constant	$0.219^{***}(0.0620)$	$0.201^{***}(0.0520)$	$0.171^{***}(0.0140)$	$0.208^{***}(0.0110)$
Observations	49,822	49,822	49,822	49,822
Log Likelihood	−1,253,579.000	-	−216,539.500	−216,655.835
θ	-	-	$0.810^{***}(0.005)$	$0.807^{***}(0.005)$
AIC	2,507,205.000		433,127.000	433,336.000

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Std. Errors in parentheses

Table 5
Explanatory models to examine the count of customer reviews from Airbnb Cluster 2.

	Dependent variable: Count of customer reviews			
	Poisson	Quasi-Poisson	Neg Bin All	Neg Bin Top10
log(Length of Summary)	$0.001^{***}(0.0030)$	$0.001^{***}(0.0030)$	$0.002^{***}(0.0022)$	
Readability	$0.006^{***}(0.0010)$	$0.006^{***}(0.0010)$	$0.005^{***}(0.0001)$	
Superlatives	$0.002^{***}(0.0001)$	$0.002^{***}(0.001)$	$0.001^{***}(0.0004)$	
Superhost	$0.441^{***}(0.0020)$	$0.440^{***}(0.0140)$	$0.377^{***}(0.0120)$	$0.424^{***}(0.0110)$
Host Identity Verified	$0.181^{***}(0.0010)$	$0.181^{***}(0.0120)$	$0.141^{***}(0.0090)$	$0.146^{***}(0.0090)$
Host has a Profile Pic	$0.250^{***}(0.0570)$	$0.271^{***}(0.0490)$	$0.156^{***}(0.0360)$	$0.310^{**}(0.0080)$
Host Duration	$0.001^{***}(0.0002)$	$0.001^{***}(0.0001)$	$0.008^{***}(0.0001)$	$0.0006^{**}(0.0007)$
Location Exact	$0.106^{***}(0.0020)$	$0.106^{***}(0.0150)$	$0.075^{***}(0.0110)$	$0.077^{***}(0.0107)$
Bed Type	$-0.278^{***}(0.0060)$	$-0.278^{***}(0.0540)$	$-0.329^{***}(0.0450)$	
log(Bedrooms)	$-0.321^{***}(0.0040)$	$-0.297^{***}(0.0170)$	$-0.306^{***}(0.0150)$	$-0.350^{***}(0.0008)$
Room Type	$-0.213^{***}(0.0020)$	$-0.213^{***}(0.0016)$	$-0.143^{***}(0.0080)$	
Amenities	$0.024^{***}(0.0085)$	$0.024^{***}(0.0085)$	$0.023^{***}(0.0040)$	$0.026^{***}(0.0001)$
log(Price per Night)	$-0.084^{***}(0.0010)$	$-0.078^{***}(0.0110)$	$-0.064^{***}(0.0080)$	
Review Scores Rating	$0.013^{***}(0.0001)$	$0.013^{***}(0.0010)$	$0.009^{***}(0.0005)$	$0.023^{***}(0.0004)$
Positive Sentiments	$0.005^{***}(0.0010)$	$0.008^{***}(0.0010)$	$0.015^{***}(0.0008)$	$0.057^{***}(0.0020)$
Negative Sentiments	$-0.012^{***}(0.0008)$	$-0.017^{***}(0.0020)$	$-0.030^{***}(0.0010)$	
Certainty	$0.020^{***}(0.0020)$	$0.020^{***}(0.0020)$	$0.062^{***}(0.0004)$	
Insights	$0.031^{***}(0.0010)$	$0.031^{***}(0.0070)$	$0.073^{***}(0.0050)$	
Instant Bookable	$0.183^{***}(0.0010)$	$0.183^{***}(0.0120)$	$0.168^{***}(0.0090)$	$0.161^{**}(0.0870)$
Cancellation Policy	$-0.180^{***}(0.0130)$	$-0.189^{***}(0.1120)$	$-0.158^{***}(0.0830)$	
Guest Phone Verification	$0.125^{***}(0.0050)$	$0.115^{***}(0.0450)$	$0.175^{***}(0.0410)$	
Availability 90	$0.002^{***}(0.0002)$	$0.002^{***}(0.0002)$	$0.003^{***}(0.0001)$	
Constant	$1.299^{***}(0.0270)$	$1.200^{***}(0.1230)$	$0.812^{***}(0.1120)$	$0.736^{***}(0.0480)$
Observations	68,380	68,380	68,380	68,380
Log Likelihood	−1,643,711.000	-	−299,145.800	−299,197.950
θ	-	-	$0.820^{***}(0.004)$	$0.819^{***}(0.004)$
AIC	3,287,469.000		598,339.600	598,419.900

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Std. Errors in parentheses.

Table 6
Explanatory models to examine the count of customer reviews from Airbnb Cluster 3.

	Dependent variable: Count of customer reviews			
	Poisson	Quasi-Poisson	Neg Bin All	Neg Bin Top10
log(Length of Summary)	0.001***(0.0002)	0.001***(0.0002)	0.001***(0.0003)	
Readability	0.004***(0.0010)	0.004***(0.0010)	0.005***(0.0018)	
Superlatives	0.001***(0.0001)	0.001***(0.001)	0.001***(0.0004)	
Superhost	0.312***(0.0020)	0.312***(0.0140)	0.319***(0.0120)	0.324***(0.0110)
Host Identity Verified	0.156***(0.0010)	0.156***(0.0010)	0.111***(0.0090)	0.164***(0.0090)
Host has a Profile Pic	0.660***(0.0150)	0.660***(0.0145)	0.467***(0.079)	
Host Duration	0.001***(0.0002)	0.001***(0.0001)	0.0005***(0.0001)	0.0005*(0.0007)
Location Exact	0.114***(0.0010)	0.115***(0.0013)	0.099***(0.0110)	0.100***(0.0107)
Bed Type	−0.181***(0.0060)	−0.181***(0.0056)	−0.226***(0.0053)	
log(Bedrooms)	−0.507***(0.0020)	−0.507***(0.0150)	−0.566***(0.0130)	−0.610***(0.0008)
Room Type	−0.211***(0.0030)	−0.211***(0.0045)	−0.188***(0.0100)	
Amenities	0.020***(0.0004)	0.020***(0.0004)	0.022***(0.0004)	0.026***(0.0001)
log(Price per Night)	−0.126***(0.0010)	−0.126***(0.0008)	−0.117***(0.0006)	
Review Scores Rating	0.014***(0.0001)	0.014***(0.0001)	0.017***(0.0050)	0.022***(0.0004)
Positive Sentiments	0.003***(0.0001)	0.003***(0.0001)	0.001***(0.0010)	
Negative Sentiments	−0.008***(0.0001)	−0.010***(0.0002)	−0.003***(0.0014)	−0.004**(0.0020)
Certainty	0.018***(0.0010)	0.018***(0.0010)	0.032***(0.0008)	
Insights	0.002***(0.0004)	0.002***(0.0003)	0.004***(0.0003)	
Instant Bookable	0.171***(0.0010)	0.171***(0.0010)	0.170***(0.0090)	0.161**(0.0870)
Cancellation Policy	−0.115***(0.0070)	−0.115***(0.0660)	−0.109***(0.0510)	
Guest Phone Verification	0.482***(0.0030)	0.482***(0.0033)	0.486***(0.0032)	
Availability 90	0.004***(0.0002)	0.004***(0.0002)	0.005***(0.0001)	0.003***(0.0001)
Constant	2.100***(0.0180)	2.100***(0.0170)	2.180***(0.0110)	2.180***(0.0480)
Observations	78,462	78,462	78,462	78,462
Log Likelihood	−2,675,474.000	-	−361,732.000	−299,197.950
θ	-	-	0.682***(0.003)	0.819***(0.004)
AIC	5,350,990.000		723,506.000	598,419.900

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Std. Errors in parentheses

Table 7
Explanatory models to examine the count of customer reviews from Airbnb Cluster 4.

	Dependent variable: Count of customer reviews			
	Poisson	Quasi-Poisson	Neg Bin All	Neg Bin Top10
log(Length of Summary)	0.002***(0.0003)	0.002***(0.0001)	0.002***(0.0001)	
Readability	0.004(0.0002)	0.004(0.0002)	0.003(0.0030)	
Superlatives	0.001*(0.0001)	0.001*(0.0004)	0.001*(0.0003)	
Superhost	0.354***(0.0020)	0.354***(0.0150)	0.391***(0.0110)	0.415***(0.0001)
Host Identity Verified	0.177***(0.0010)	0.177***(0.0120)	0.075***(0.0080)	0.079**(0.0080)
Host has a Profile Pic	0.132**(0.0120)	0.132**(0.0140)	0.131**(0.0100)	
Host Duration	0.002***(0.0001)	0.002***(0.0001)	0.006***(0.0001)	0.006**(0.0001)
Location Exact	0.097***(0.0020)	0.097***(0.0120)	0.038***(0.0080)	0.041**(0.0080)
Bed Type	−0.239***(0.0060)	−0.238***(0.0510)	−0.193***(0.0430)	
log(Bedrooms)	−0.371***(0.0130)	−0.371***(0.0190)	−0.395***(0.0120)	−0.404***(0.0020)
Room Type	−0.013***(0.0010)	−0.013***(0.0011)	−0.008***(0.0050)	
Amenities	0.021***(0.0001)	0.021***(0.0001)	0.018***(0.0004)	0.020***(0.0004)
log(Price per Night)	−0.255***(0.0090)	−0.265***(0.0085)	−0.227***(0.0060)	−0.278***(0.0060)
Review Scores Rating	0.006***(0.0001)	0.006***(0.0010)	0.009***(0.0004)	
Positive Sentiments	0.006***(0.0002)	0.006***(0.0010)	0.011***(0.0010)	
Negative Sentiments	−0.004***(0.0040)	−0.003***(0.0010)	−0.012***(0.0010)	
Certainty	0.007***(0.0004)	0.007***(0.0030)	0.010***(0.0020)	
Insights	0.019***(0.0010)	0.019***(0.0080)	0.021***(0.0050)	
Instant Bookable	0.154***(0.0020)	0.154***(0.0014)	0.162***(0.0080)	0.173***(0.0011)
Cancellation Policy	−0.204***(0.0020)	−0.204***(0.0110)	−0.082***(0.0070)	−0.158***(0.0065)
Guest Phone Verification	0.526***(0.0040)	0.526***(0.0031)	0.388***(0.0030)	0.406***(0.0030)
Availability 90	0.001***(0.0002)	0.001***(0.0002)	0.001***(0.0001)	
Constant	3.461***(0.0160)	3.461***(0.0127)	3.294***(0.0240)	2.744***(0.030)
Observations	85,068	85,068	85,068	85,068
Log Likelihood	−1,652,053.000	-	−346,946.600	−348,328.900
θ	-	-	0.861***(0.0038)	0.839***(0.004)
AIC	3,304,153.000		693,917.000	696,705.800

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Std. Errors in parentheses

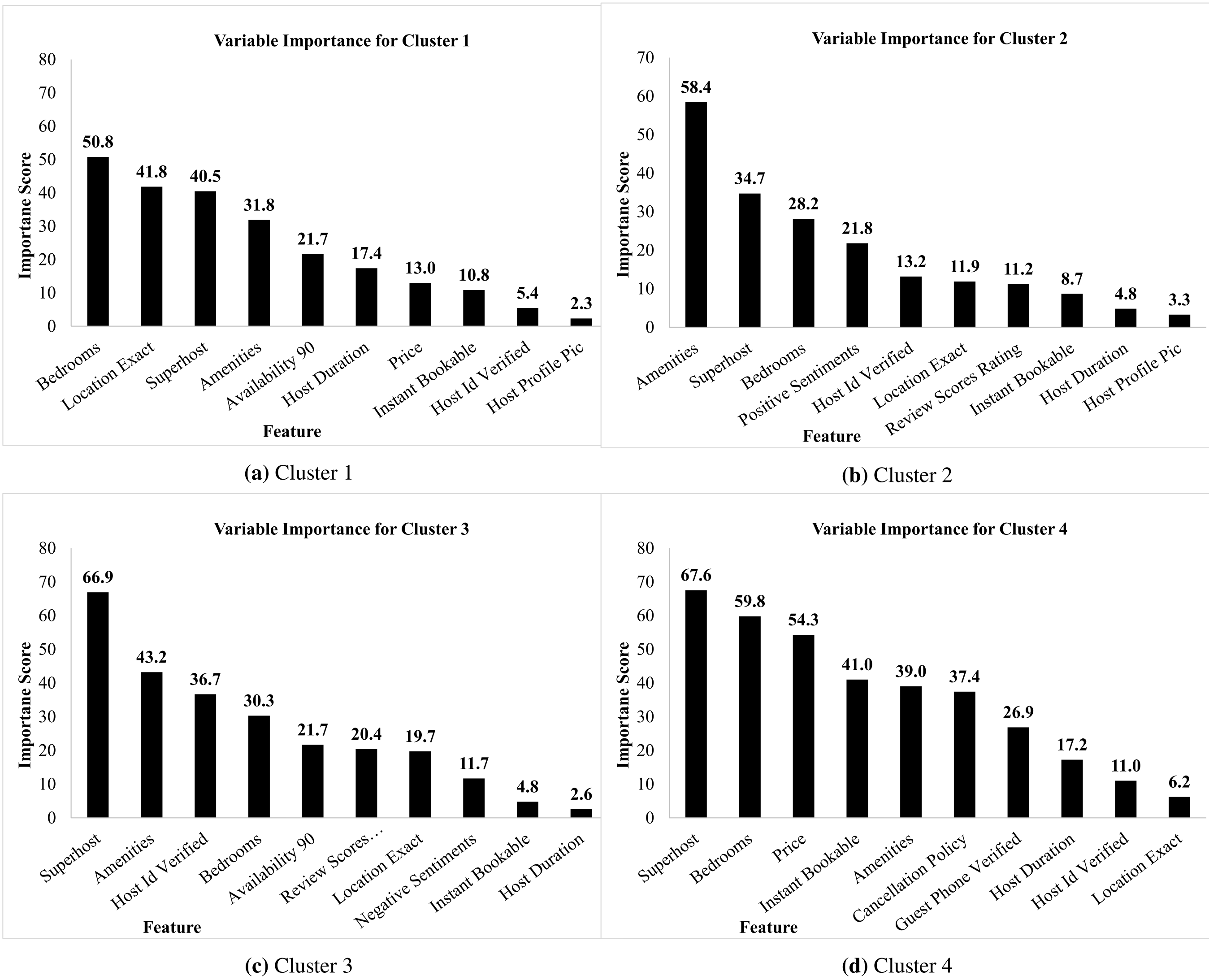


Fig. 7. Top 10 features across 4 clusters using importance scores

variables, such as the *review certainty*, *positive sentiments*, and *90-day availability*.

5. Discussion of results

5.1. Main effects of empirical analysis across clusters

We present the primary effects of the determinants and their related findings that can help hosts understand how to increase the number of reviews received (as a consequence of successful bookings) through the improvement of the following six dimensions of their accommodations, namely, *website, host, property, historical reviews, rental policies, and availability*.

First, we found that the *website* attributes have a positive effect on the count of customer reviews received by Airbnb properties. The more detailed a self-description is, the better off everyone would be since each Airbnb host is different.[14] Liang et al. [14] have identified that providing comprehensive and detailed descriptions related to Airbnb properties can lead to more reviews. Airbnb also recommends that a detailed and comprehensive description of the property, added with an alluring title, can attract more bookings and lead to a higher count of customer reviews received. Next, we found that the coefficients for the emotiveness of the property description as posted by the host (MGC) are similar to those reported by Siering et al. [29]. When hosts use more adjectives, such as *a perfect room,* or *airy space*, the property is perceived to generate more value for the guests. When compared across clusters, we found that irrespective of the constituent cities, guests considered the *website* attributes as "hygiene-factors," and therefore, their coefficients estimates were consistently positive, statistically significant, and values were close. Throughout our analysis, we purposefully omitted the examination of the *sentiments of self-description texts* as a determinant because most hosts would usually write positive words about themselves without adding critical reviews and therefore have little(or no) business value.

Second, we found that the *host attributes* in our study acted as "reputation systems" and helped to create a positive and significant effect on the count of customer reviews that were generated from successful bookings. The slogan of Airbnb is to "belong anywhere."[15] Thus, guests prefer to associate themselves with the home of a trustworthy Airbnb

[14] How to Be the Perfect Airbnb Host (or a Superhost) | HuffPost Life, Retrieved From: https://www.huffpost.com/entry/how-to-be-the-perfect-air_b_8539700

[15] The Airbnb Story: Belong Anywhere, Retrieved From: https://blog.atairbnb.com/belong-anywhere/

Table 8
Using "Superhost" as a moderator to examine the count of customer reviews.

	Dependent variable: Count of customer reviews			
	NB-C1	NB-C2	NB-C3	NB-C4
Readability	0.009*(0.0020)	0.007*(0.0019)	0.003*(0.0030)	0.008*(0.0023)
Superlatives	0.005*(0.0006)	0.002*(0.0004)	0.001*(0.0005)	0.001*(0.0003)
Location Exact	0.164***(0.0130)	0.067***(0.0120)	0.054***(0.0080)	0.138**(0.0090)
Bed Type	−0.383***(0.0650)	−0.295***(0.0500)	−0.162***(0.0120)	−0.232***(0.0490)
log(Bedrooms)	−0.675***(0.0170)	−0.313***(0.0160)	−0.404***(0.0190)	−0.400***(0.0140)
Room Type	−0.052***(0.0140)	−0.102***(0.0110)	−0.096***(0.0160)	−0.065***(0.0080)
Amenities	0.015***(0.0005)	0.025***(0.0005)	0.021***(0.0004)	0.019***(0.0040)
log(Price per Night)	−0.092***(0.0080)	−0.020***(0.0080)	−0.023***(0.0070)	−0.144***(0.006)
Superhost	0.739***(0.0140)	0.585***(0.0137)	0.761***(0.0138)	0.516***(0.0122)
Superhost * Superlatives	0.002*(0.0010)	0.003*(0.0010)	0.001*(0.0010)	0.0005*(0.0001)
Superhost * Readability	0.002*(0.0010)	0.005*(0.0020)	0.001*(0.0003)	0.001*(0.0025)
Superhost * Location Exact	0.167***(0.0260)	0.073***(0.0250)	0.060***(0.0100)	0.133**(0.0023)
Superhost * Bed Type	−0.077**(0.0320)	−0.017***(0.0118)	−0.043***(0.0600)	−0.056**(0.0106)
Superhost * log(Bedrooms)	−0.075**(0.0310)	−0.059***(0.0330)	−0.097***(0.0310)	−0.138***(0.0400)
Superhost * Room Type	−0.010***(0.0010)	−0.048***(0.0010)	−0.038***(0.0030)	−0.015***(0.0040)
Superhost * Amenities	0.009***(0.0010)	0.020***(0.0010)	0.018***(0.0010)	0.015***(0.0010)
Superhost * log(Price per Night)	−0.136***(0.014)	−0.047***(0.0170)	−0.054***(0.0160)	−0.185***(0.0017)
Review Scores Rating	0.010***(0.0007)	0.020***(0.0005)	0.013***(0.0005)	0.005***(0.0004)
Positive Sentiments	0.009***(0.0008)	0.020***(0.0010)	0.010***(0.0010)	0.010***(0.0010)
Negative Sentiments	−0.007***(0.0050)	−0.042***(0.0020)	−0.032***(0.0010)	−0.030***(0.0009)
Certainty	0.051***(0.0030)	0.062***(0.0020)	0.064***(0.0020)	0.011***(0.0020)
Insights	0.089***(0.0071)	0.072***(0.0050)	0.038***(0.0050)	0.017***(0.0020)
Constant	2.846***(0.0930)	0.842***(0.0740)	2.800***(0.0790)	3.223***(0.064)
Observations	49,822	68,380	78,462	85,068
Log Likelihood	−216,640.500	−299,348.800	−361,930.500	−348,843.400
θ	0.808***(0.005)	0.816***(0.004)	0.679***(0.003)	0.830***(0.004)
AIC	433,328.900	598,739.600	433,195.000	697,728.900

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Std. Errors in parentheses.
NB-Ci = Negative Binomial with Cluster *Ci* data, where i = 1,2,3,4.

Table 9
Sensitivity analysis for our explanatory models (t = Jan 2014 up to Jan 2020).

	Dependent variable: Count of customer reviews			
	NB-C1	NB-C2	NB-C3	NB-C4
log(Length of Summary)	0.0003**(0.0002)	0.004**(0.0002)	0.001**(0.0002)	0.001*(0.0002)
Readability	0.068**(0.0023)	0.081**(0.0021)	0.092**(0.0019)	0.012**(0.0180)
Superlatives	0.004***(0.0010)	0.004***(0.0010)	0.005***(0.0010)	0.004***(0.0003)
Superhost	0.550***(0.0150)	0.381***(0.0120)	0.368***(0.0130)	0.493***(0.0130)
Host Identity Verified	0.033**(0.0100)	0.035**(0.0120)	0.027**(0.0090)	0.060***(0.0090)
Host has a Profile Pic	0.038**(0.0970)	0.270**(0.0002)	0.347**(0.0008)	0.150**(0.0074)
Host Duration	0.009***(0.0004)	0.006***(0.0002)	0.006***(0.0001)	0.008***(0.0003)
Location Exact	0.132**(0.0130)	0.173**(0.0120)	0.152**(0.0120)	0.024**(0.0090)
Bed Type	−0.390***(0.0702)	−0.319***(0.0550)	−0.144**(0.0590)	−0.156***(0.0550)
log(Bedrooms)	−0.579***(0.0160)	−0.592***(0.0150)	−0.621***(0.0140)	−0.351***(0.0140)
Room Type	−0.004*(0.0080)	−0.011***(0.0030)	−0.012***(0.0070)	−0.023***(0.003)
Amenities	0.032***(0.0010)	0.013***(0.0010)	0.012***(0.0005)	0.011***(0.0001)
log(Price per Night)	−0.091***(0.0070)	−0.075***(0.0110)	−0.106***(0.0070)	−0.289***(0.0060)
Review Scores Rating	0.012***(0.0010)	0.020***(0.0005)	0.019***(0.0004)	0.007***(0.0004)
Positive Sentiments	0.012***(0.0010)	0.008***(0.0010)	0.008***(0.0010)	0.013***(0.0010)
Negative Sentiments	−0.053***(0.0350)	−0.085***(0.0460)	−0.059***(0.0010)	−0.039***(0.0140)
Certainty	0.047***(0.0030)	0.041***(0.0030)	0.041***(0.0020)	0.006**(0.0030)
Insights	0.063***(0.0015)	0.068***(0.0070)	0.086***(0.0070)	0.010**(0.0060)
Instant Bookable	0.125***(0.0120)	0.159***(0.0110)	0.150***(0.0110)	0.308***(0.0090)
Cancellation Policy	−0.295***(0.0220)	−0.247***(0.0220)	−0.206***(0.0200)	−0.244***(0.0090)
Guest Phone Verifi cation	0.185**(0.0090)	0.076*(0.0370)	0.035*(0.0360)	0.047*(0.0290)
Availability 90	0.013***(0.0002)	0.003***(0.0002)	0.003***(0.0001)	0.001***(0.0003)
Constant	1.237***(0.0207)	1.343***(0.0123)	1.464***(0.0120)	2.486***(0.0103)
Observations	36,305	52,364	59,168	64,114
Log Likelihood	−154,141.306	−226,354.700	−269,294.900	−256,979.200
θ	0.846***(0.006)	0.842***(0.005)	0.691***(0.004)	0.877***(0.005)
AIC	308,326.600	452,751.500	538,631.800	514,000.300

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Std. Errors in parentheses
NB-Ci = Negative Binomial with Cluster *Ci* data, where i = 1,2,3,4

Table 10
A simplified actionable scorecard to improve the number of reviews on Airbnb

	Website	Host	Property	Historical Review	Rental Policy	Availability
Cluster 1	No	4	4	No	1	1
Cluster 2	No	4	3	2	1	No
Cluster 3	No	3	3	2	1	1
Cluster 4	No	3	4	No	3	No

*Ranking among the top 10 features

host, and so the "Superhost" badge is of utmost importance to them [14,20,34]. Next, we found that if the hosts had their identity documents verified or could present trust-related cues, for instance, included a profile photo [22]; then they were more likely to attract customers due to host trustworthiness [21,23,26]. Then, we found that the *host duration in months* positively affected the number of reviews received, since older properties signalled a higher level of trust to the customers [24].

Third, we examined the effects of various *property-based attributes* on the count of customer reviews received by those properties as a consequence of successful Airbnb bookings. We noted that the *exact location* of the property had a positive effect, which has also been reported as a significant predictor in the literature [28,52]. London (Cluster 4) has the lowest coefficient estimate because compared to other cluster-cities, it is much bigger and possesses an excellent public transportation network. Hence, even if the Airbnb property is not at the exact location as promised, visitors might not face any difficulties in reaching it. Again, in Sydney (Cluster 3), not all listings are close to the city centre, and visitors might need a car to get around. In contrast, Brussels (Cluster 1) and Sydney (Cluster 3) attract large vacationing crowds while London draws more business visitors. Next, we found that the effect of *bed-type* was negative because the guests preferred a real bed instead of an airbed, futon, or a sofa. Coefficient estimates were the highest for Cluster 1 and lowest for Cluster 4, indicating that visitors staying in London (Cluster 4) could compromise over a real bed and still rent Airbnb properties, but not travellers to "Cluster 1" cities. This result, hitherto unreported in the relevant literature, also implies that the *bed-type* offered at Airbnb properties could affect the subsequent count of customer reviews.

Next, we found that the *number of bedrooms* had a significant negative effect towards the count of reviews received by Airbnb properties across all clusters. If a higher number of bedrooms are listed for a property, the chances are that distinct travellers might rent them. Guests could perceive this action as a compromise of their privacy and hence avoid reservations. The result is also in congruence with our findings from the room-type variable, where we had found that staying at an entire apartment was preferred to the shared rooms. While Xie and Mao [56] reported a positive but insignificant effect of the *number of bedrooms* on the Airbnb reservations, the results from our study are consistent with Liang et al. [14]. The latter also found a significant and negative effect of the *number of bedrooms* on the *number of reviews* received.

We also examined the effects of *room type* on the count of customer reviews. Room-type was defined as a categorical variable with an entire apartment as the benchmark (please see Table 1). Results from our study indicated that the coefficient estimates were lowest for Cluster 4 and highest for Cluster 1. To some extent, this can be explained by the fact that London (Cluster 4) is a global business hub, and therefore, a *private room* for business travellers would suffice. However, for other cluster-cities, family travellers might prefer a bigger place with more privacy. Thus, staying at an entire apartment would serve their purpose more than any other option. Qiu et al. [36] found that an entire property had a 290.84 % higher chance of being reserved than shared rooms. Extant literature has also indicated the importance of *room type* as an effective determinant of the number of customer reviews and a willingness to pay more price [14,60].

We then examined the effect of the *number of unique amenities offered* on the count of customer reviews received by the Airbnb property. We found that irrespective of cluster-cities, guests considered amenities as "hygiene factors," and therefore, their coefficients estimates were consistently positive, statistically significant, and values were close. According to scholars [10,35], the Airbnb properties with more unique amenities, such as wireless Internet, free parking, kitchen, and laundry services, could easily increase their customer appeal as compared to those without such facilities. Next, we found a consistent and significant negative effect of *price per night* across all four clusters. Cluster 4 (London) exhibits the highest negative coefficient for per unit change of the number of reviews with *price per night*, everything else remaining constant. Marshall's classical economic theory [61] states that the price of a good is based on both the demand and supply. Therefore, in the case of Cluster 4 (London), which has the highest number of Airbnb properties, a highly competitive pricing mechanism exists as compared to others. However, for the remaining cluster-cities where the count of properties is low, guests might perceive the listed price as a signal of quality of the good (i.e., Airbnb property). Therefore, a relatively weaker but negative relationship exists for clusters 1, 2, and 3. Our results are also congruent with Qiu et al. [36], who noted that an increase in the price per night by one pound sterling would decrease the possibility of reservations by 1.16%.

Fourth, we examined the effects of the characteristics of historical reviews received by an Airbnb property, namely, *review scores ratings*, *average positive sentiments*, *average negative sentiments*, *cognitive certainty*, and *cognitive insights*. The coefficient estimates for *review scores ratings* are consistently positive, statistically significant, and numerically close across clusters, signifying that they have a mandatory effect irrespective of cluster-cities. However, Cluster 4 shows the lowest estimate among them. The potential reason could be the huge number of listings available in Cluster 4, and therefore guests are less concerned with the ratings. Our finding also matches the mean rating of each cluster, which is the lowest for Cluster 4 and the highest for Cluster 1. Next, we found that the coefficients for *average positive sentiments* and *average negative sentiments* were the highest for Cluster 2 and lowest for Cluster 3. An interesting observation was that the coefficients for the negative sentiments were almost twice the positive sentiments on clusters 2 and 3. Liang et al. [14,20] have also reported similar findings. Next, we found that both *cognitive certainty* and *cognitive insights* generated positive impacts on the number of customer reviews received. This finding is in congruence with our proposed analytical framework. We noted that these determinants had a much stronger effect for Clusters 1 and 2 than for Clusters 3 and 4. One possible reason could be due to the high number of properties listed in Clusters 3 and 4. Therefore, prospective guests did not pay much attention to the detailed historical reviews received by each property in these clusters, resulting in weak cognitive emotions.

Fifth, we examined the effects of *rental policies and rules* on the count of customer reviews. We defined *cancellation policy* as a categorical variable with *strict cancellation* as the benchmark category. In case of a hotel reservation, standard cancellation policies apply, whereas flexible and friendly cancellation policies for Airbnb rentals can render the "ease of choice" for guests to cancel if necessary. We found that the coefficient estimates were much lower for Clusters 3 (Los Angeles and Sydney) and 4 (London), and relatively higher for the rest. These cluster-cities, being business centres, attract many travellers who plan last minute or on arrival. Therefore, hosts have maintained flexible rental policies to accommodate these guests. On the contrary, Clusters 1 and 2 have a large pool of tourists to cater to, who plan well ahead. Hence a stricter cancellation rule applies to these clusters. Next, we found that the coefficients of *guest phone verification* had a consistent, significant, and positive impact on the count of customer reviews. While Clusters 1 and 2 had a lower influence of guest phone verification, Clusters 3 and 4 had a relatively higher impact. This effect could be due to the massive arrival of multinational guests within a short time, and hosts need to maintain the security of their properties.

Finally, we examined the effects of the *90-day availability* on the count of customer reviews, which showed a positive impact across all four clusters. While extant literature on Airbnb had considered the availability calendar for 365 days in the future and its effect on the prospective guests' behaviour, such treatment fails to represent a realistic vacation-planning process. When prospective guests plan for a vacation or short-term rental, they usually search within a span of the coming 3 months. Considering a 90-day period also has a governance perspective associated with it. Airbnb has put a limit on the number of nights that hosts can keep their homes open for reservations. Also known as the *90-Day Rule*, it was implemented to legalize short-term P2P shared-home rentals within the Greater London Area. [16]

Using the variable importance scheme (column "Neg Bin Top10" from Tables 4, 5, 6, and 7) and Fig. 7, we build Table 10 to present a simplified yet actionable scorecard for Airbnb hosts who can work on these characteristics as potential areas for improvement to enhance the quality of their accommodations and thereby attract more customer reviews.

5.2. Trust as a moderator of MGC effects on the count of customer reviews

Among the available host attributes, "superhost" badge acts as the most suitable and encompassing trust-building mechanism and reputation system for Airbnb hosts. We reported several interesting findings from the moderator effects of "superhost" between marketer-generated-content (MGC-s) and the count of customer reviews. We found that the "superhost" status could boost the effects of the determinants while guests stayed at Airbnb homes. For example, the "superhost" badge enhanced the effects of "bed type" and "room type," indicating that properties with a "superhost" badge had lesser chances of multiple bedrooms on offer. Therefore, they could allocate more of "entire homes" and "full rooms" than "shared rooms," thereby granting more privacy to the incoming guests. This scenario enhanced the privacy concerns of the guests who had already reserved a particular Airbnb property. We also found that the guests were willing to spend extra for staying on properties with fewer bedrooms or more beds.

On the contrary, we found that guests actually paid a higher "price per night" for properties owned by "superhosts." While classical economic theory [61] suggests that an adverse effect of "price per night" on the count of customer reviews is expected, this reverse mechanism can further intensify in the presence of the "superhost" badge. However, the phenomenon is easily explained by the social exchange theory [17], which suggests that "superhosts" invest their accumulated "reputational" capital to gain the trust of the guests, and thereby use it to set higher property prices. These factors enable the niche group of 19.4% "superhosts" to earn 60% more revenue than other regular hosts worldwide.[17] Next, we found that if a "superhost" maintained the property, guests happily overlooked necessary "website attributes" namely, (i) the adjectives used to describe a property, such as *a perfect room*, or *airy space*; and (ii) readability of self-description of a property, each of which would otherwise be deemed essential. Finally, we found that guests considered the "amenities" offered at the property and its "exact location" to be "hygiene factor"-s. Therefore the moderator effect of the "superhost" badge on these determinants was found to be negligible for the count of customer reviews.

[16] 90 day rule for London Airbnbs, Retrieved From: https://www.airbnb.co.in/help/article/1340/i-rent-out-my-home-in-london-what-shortterm-rental-laws-apply

[17] https://www.airdna.co/blog/airbnb_superhost_status

6. Implications of our study

6.1. Theoretical implications

Our study has several theoretical contributions, mainly towards the social penetration theory [46], social exchange theory [17], and value co-creation theory [12] in the context of the flourishing P2P shared-home rental industry. First, our study extends the following three aspects of social penetration theory - (1) identifies how the interpersonal relationships between a guest and a host develop - before, during, and after a P2P shared-home experience; (2) proposes innovative *implicit* interactions between the guest and the host which are represented by the linguistic cues that were derived from the eWOMs (i.e., guest reviews). These determinants are *positive* and *negative* review sentiments, cognitive dimensions such as *insights*, and *certainty*; (3) while P2P shared-homes require robust trust-building mechanisms and two-way safety verifications to encourage trustworthiness between the guest and the host, our study demonstrated that the *superhost* feature, *profile photos*, *verified host identity*, *membership duration*, and *guest verification* could successfully build "trust" in the ongoing interpersonal relationships during a collaborative consumption.

Second, our study contributes to the following three aspects of social exchange theory - (1) integrate a set of robust and unique features into the ensuing cost-benefit analysis through a collaborative consumption after the "social exchange." Primarily these were UGC features such as *overall average review ratings received*, *review sentiments*, and *cognitive emotional contents of reviews* as well as MGC features, such as *text readability of the property description*, *superlatives used in the title of the property description*, the *exact location of the property*, and *cancellation policy*; (2) while hosts invested their accumulated "reputational" capital (such as *superhost* tag, *verified identity*, and *longer membership duration*) in charging a rental price, the guests expected enjoyment (*good amenities, real beds*, and *private rooms*); (3) evidence of "price-rise" among "superhost" properties, who reaped their status to accrue the trust of guests and thereby used it to charge higher property prices which guests were often willing to pay.

Third, our study contributes to the following three aspects of value co-creation theory - (1) create "value" that is determined by the guest experience (and received in the form of eWOMs) throughout the collaborative consumption process, rather than considering "value" as pre-existing; (2) successful "value co-creation" through the enhancement of both UGC and MGC characteristics of P2P accommodations, such as a property with good amenities, an attractive and detailed website listing, flexible cancellation policies; (3) the online trust achieved in the form of a *superhost badge*, *verified host identity*, *host's profile photo*, and *verified guest identity* is a unique-valued asset that is important for both the hosts and the guests because it encourages "value co-creation."

Fourth, we introduced a novel listing-based clustering mechanism for the Airbnb platform during our empirical investigation. While previous studies had empirically analyzed the aggregated count data collected from all locations together, we extended them by segregating the data into homogeneous clusters, based on the count of properties in each city. This mechanism helped us to highlight the cluster-specific variations and examine the effects of those determinants separately on the count of customer reviews received by Airbnb properties. Therefore, we propose that a *one-size-fits-all* strategy cannot be adopted during an examination of these determinants because it might yield non-actionable recommendations for the hosts residing in these cities. Thus, our study extends the applications of the bid-rent theory [62], transaction cost, and information search economics [3], and firm agglomeration theory [63] in the context of the P2P shared-home rentals.

Fifth, from the variable importance schemes (Fig. 7), we identified that the *price per night* is a significant feature for Cluster 4 only, but is ranked much below among the determinants across other cluster-cities. While the high price of an economic good (i.e., P2P shared-home) in the rental market may signal a level of higher quality among its competitors and vice versa, merely choosing competitive pricing to showcase them may lead to a *market for lemons and oranges* [64]. Gradually, good quality shared properties may remain unreserved, and their hosts might leave the shared-home rental industry due to poor (or no) profitability, leaving only a pool of mediocre listings. While such an "adverse selection problem" has not been utterly unseen in the hospitality industry as in the case of Oyo [18], we believe that our study can guide Airbnb hosts to eradicate the incumbent price-quality interdependence behaviour in the P2P shared-home rental market through continuous improvements of non-price characteristics, such as amenities, detailed and legible self-descriptions on the website, and finally, guest-friendly rental and cancellation policies.

6.2. Managerial implications

Our study is also essential for managers due to several reasons. First, it gives an idea about how to forecast the number of customer reviews received by P2P property reservations. This finding, we believe, would also help the hosts to take improvement measures and highlight various characteristics of their listings to potential guests who search for peer-to-peer shared-homes online. On the other hand, a consumer can also make improved and informed decisions if they are fully aware of the factors to look for while searching for a P2P shared-home and thus help both the parties in better decision-making, thereby leading to a post-purchase satisfaction.

Second, with the exponential growth in the P2P shared-home rental industry, new businesses are emerging to assist Airbnb hosts who are continuously looking to improve the performance of their short-term rentals. Among many, Evolve and Turnkey in North America, and Guest-Ready in Europe are some successful property management firms that handle P2P shared-home rentals [19]. In this context, our study offers a set of "cluster-based" actionable recommendations, which we believe brings in more customer reviews for the hosts and therefore appeal to prospective guests.

Third, we found that an array of features emerged as the significant attributes of P2P shared-home rentals apart from *price per night*. The variable importance scheme found that the *price per night* ranked third for Cluster 4 (London) only, and did not appear among the top 5 features for the remaining clusters. This finding is in contrast to the existing corpus of literature on P2P sharing-economy based accommodations that have examined the determinants of Airbnb pricing. It also helps to corroborate our results with the recommendations that Airbnb suggests for its hosts [20]. While Airbnb advocates that hosts may set competitive pricing, they also suggest that improving *self-description of the property*, setting *flexible rental policies*, and an alluring *title* can make the property competitive and lead to better review ratings.

Fourth, we have shown that, along with the MGC features [14], the UGC-based covariates can also generate positive outcomes from customers, more so in the case of shared-home rentals. This finding has significant implications for business managers while designing unique mobile-based marketing strategies for tourists and service highlights for shared-property owners. Accordingly, the advertisement contents displayed on various media for shared-home rental firms such as Airbnb, Guesthouser, and Stayaway can be efficiently designed to suit prospective guests as well as hosts.

[18] India's OYO Needs to Focus on Hotel Quality, Not Just Expansion, Retrieved From: https://www.bloomberg.com/opinion/articles/2018-09-26/india-s-oyo-needs-to-focus-on-hotel-quality-not-just-expansion

[19] AirDNA:The Best Airbnb Property Management Services, Retrieved From: https://www.airdna.co/blog/the-best-airbnb-property-management-services

[20] How do I make my listing more competitive? Retrieved From: https://www.airbnb.co.in/help/article/431/how-do-i-make-my-listing-more-competitive

7. Conclusion and future scope

In this study, we identified the determinants of the number of customer reviews received by P2P shared-homes using 22 independent features that were categorized as: *website*, *host*, *property*, *historical reviews*, *rental and cancellation policies*, and *90-day availability*. We considered the "count of customer reviews" as a proxy measure for successful bookings. We performed a cluster analysis based on the listing numbers to generate homogenous "cluster cities." We then analyzed the Airbnb properties using Poisson regression and subsequently with Quasi-Poisson and Negative Binomial regression techniques due to the presence of overdispersion effects. We further accompanied the count regressions with variable importance schemes and identified the top 10 features separately for each cluster. We also investigated the effect of "reputation systems" on the host-generated features towards the count of customer reviews through an examination of the moderation effects of the "superhost" badge on their MGC features. We further noted that the effects of each of these predictors appeared with similar signs and comparable estimates across the four clusters, thereby indicating homogeneity and robustness in predicting the count of customer reviews using our proposed analytical framework ([14,43]. Therefore, our study is among the first to identify, examine, and compare the determinants of customer reviews of the P2P shared-home rentals using data from Airbnb properties. Although we do not claim to be fully inclusive, the findings from our study offer insights into the complexities of the relationships and the importance of these determinants, which we believe shall benefit both the hosts and the guests to plan a successful collaborative consumption experience.

Specifically, this study identified the importance of several host-generated and user-generated features across the four clusters. Among host-generated content, *superhost badge*, *host duration*, *bedrooms*, and *amenities offered* were significant. Among user-generated content, *overall review scores*, and *negative sentiments* were significant. We found that the "superhost" badge moderated the effects of host-generated features on the count of customer reviews received by the shared-home. Therefore, the guests were willing to pay a higher "price per night" for the "superhost" properties. However, they ignored a few features while searching for these "superhost" properties, for instance, excellent self-descriptions on the website. We noted that the guests considered the "amenities" being offered at the property and its "exact location" to be "hygiene factor"-s and therefore, the moderating effect of the "superhost" badge on these determinants was found to be negligible. Based on the variable importance scheme, we built a simple yet actionable scorecard as a ready recommendation tool for the hosts on the Airbnb platform to improve the number of customer reviews that they subsequently receive. Thus, we proposed that a "cluster-specific" approach needs to be adopted for the Airbnb stakeholders instead of a *one-size-fits-all* strategy.

Despite all these unique contributions, our study is not free from limitations. First, we worked with cross-sectional property-based data extracted from Airbnb to examine the determinants. In the future, advanced econometric models can use longitudinal data to incorporate seasonality, festivals, and other socio-economically relevant events. Second, in this study, we identified that Airbnb primarily relies on the mutual trust between a host and a guest to generate shared-home reservations. Thus, a detailed analysis of the trust-building factors using profile-photo features, gender, and race [65,66] can be considered as a consequential extension of this study. Third, the property-based data collected from the ten cities in our study did not have any unreserved homes. Hence, we limited our modelling to Poisson, Quasi-Poisson, and Negative Binomial regressions. If the need arises, future studies can incorporate zero-inflated distributions to handle the overdispersion effects due to a large proportion of unutilized shared properties.

Acknowledgements

The authors wish to express their sincerest thanks to the anonymous referees for their encouraging suggestions that led to significant improvements to this paper. The authors also wish to thank the Editor-in-Chief of this journal, Professor James R. Marsden, for his overall inputs, continued guidance, and support throughout the revision process.

References

[1] R. Belk, You are what you can access: Sharing and collaborative consumption online, J. Bus. Res. 67 (8) (2014) 1595–1600.
[2] B. Jiang, L. Tian, Collaborative consumption: Strategic and economic implications of product sharing, Manag. Sci. 64 (3) (2018) 1171–1188.
[3] X. Xu, How do consumers in the sharing economy value sharing? evidence from online reviews, Decis. Support. Syst. 113162 (2019).
[4] L. Gansky, The mesh: Why the future of business is sharing, Penguin, 2010.
[5] A. Sundararajan, From zipcar to the sharing economy, Harv. Bus. Rev. 1 (2013).
[6] R. Botsman, R. Rogers, What's Mine Is Yours. *The Rise of Collaborative Consumption*, Harper Collins, New York, 2010.
[7] V. Katz, Regulating the sharing economy, Berkeley Technol. Law J. 30 (4) (2015) 1067–1126.
[8] G. Zervas, D. Proserpio, J.W. Byers, The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry, J. Mark. Res. 54 (5) (2017) 687–705.
[9] C. Koopman, M. Mitchell, A. Thierer, The sharing economy and consumer protection regulation: The case for policy change, J. Bus. Entrepreneurship L 8 (2014) 529.
[10] D. Guttentag, Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector, Curr. Issue Tour. 18 (12) (2015) 1192–1217.
[11] I. Blal, M. Singal, J. Templin, Airbnb's effect on hotel sales growth, Int. J. Hosp. Manag. 73 (2018) 85–92.
[12] S.L. Vargo, R.F. Lusch, Evolving to a new dominant logic for marketing, J. Mark. (2004) 1–17.
[13] L. Zhang, Q. Yan, L. Zhang, A computational framework for understanding antecedents of guests' perceived trust towards hosts on Airbnb, Decis. Support. Syst. 115 (2018) 105–116.
[14] S. Liang, M. Schuckert, R. Law, C.-C. Chen, The importance of marketer-generated content to peer-to-peer property rental platforms: Evidence from Airbnb, Int. J. Hosp. Manag. 84 (2020) 102329.
[15] S.M. Mudambi, D. Schuff, Research note: What makes a helpful online review? a study of customer reviews on amazon. com, MIS Q. (2010) 185–200.
[16] M. Salehan, D.J. Kim, Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics, Decis. Support. Syst. 81 (2016) 30–40.
[17] R.M. Emerson, Social exchange theory, Annu. Rev. Sociol. 2 (1) (1976) 335–362.
[18] D. Wang, J.L. Nicolau, Price determinants of sharing economy based accommodation rental:A study of listings from 33 cities on Airbnb.com, Int. J. Hosp. Manag. 62 (2017) 120–131.
[19] J. Oskam, A. Boswijk, Airbnb: the future of networked hospitality businesses, J. Tourism Futures 2 (1) (2016) 22–42.
[20] S. Liang, M. Schuckert, R. Law, C.-C. Chen, Be a "superhost": The importance of badge systems for peer-to-peer rental accommodations, Tour. Manag. 60 (2017) 454–465.
[21] E. Ert, A. Fleischer, N. Magen, Trust and reputation in the sharing economy: The role of personal photos in Airbnb, Tour. Manag. 55 (2016) 62–73.
[22] A. Fagerstrøm, S. Pawar, V. Sigurdsson, G.R. Foxall, M. Yani-de Soriano, That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on Airbnb™, Comput. Hum. Behav. 72 (2017) 123–131.
[23] I.P. Tussyadiah, Strategic self-presentation in the sharing economy: Implications for host branding, Information and Communication Technologies in Tourism 2016, Springer, 2016, pp. 695–708.
[24] K. Xie, Z. Mao, The impacts of quality and quantity attributes of Airbnb hosts on listing performance, Int. J. Contemp. Hosp. Manag. 29 (9) (2017) 2240–2260.
[25] E. Goffman, et al., The presentation of self in everyday life, Harmondsworth, London, 1978.
[26] X. Ma, J.T. Hancock, K. Lim Mingjie, M. Naaman, Self-disclosure and perceived trustworthiness of airbnb host profiles, Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2017, pp. 2397–2409.
[27] D. Dann, T. Teubner, C. Weinhardt, Poster child and guinea pig–insights from a structured literature review on Airbnb, Int. J. Contemp. Hosp. Manag. 31 (1) (2019) 427–473.
[28] I. Tussyadiah, F. Zach, Hotels vs. peer-to-peer accommodation rentals: Text analytics of consumer reviews in portland, oregon, Peer-to-Peer Accommodation Rentals: Text Analytics of Consumer Reviews in Portland, Oregon, April 10, 2015.
[29] M. Siering, J.-A. Koch, A.V. Deokar, Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts, J. Manag. Inf. Syst. 33 (2) (2016) 421–455.
[30] A. Bilgihan, M. Bujisic, The effect of website features in online relationship marketing: A case of online hotel booking, Electron. Commer. Res. Appl. 14 (4) (2015) 222–232.
[31] A.G. Mauri, R. Minazzi, M. Nieto-Garca, G. Viglia, Humanize your business. the role of personal reputation in the sharing economy, Int. J. Hosp. Manag. 73 (2018) 36–43.
[32] C.-F. Chen, R. Rothschild, An Application of Hedonic Pricing Analysis to the Case of Hotel Rooms in Taipei, Tour. Econ. 16 (3) (2010) 685–694.
[33] U. Gunter, What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area, Tour. Manag. 66 (2018) 26–37.
[34] E. Ert, A. Fleischer, The evolution of trust in Airbnb: A case of home rental, Ann. Tour. Res. 75 (2019) 279–287.
[35] M. Chattopadhyay, S.K. Mitra, Do airbnb host listing attributes influence room pricing homogenously? Int. J. Hosp. Manag. 81 (2019) 54–64.
[36] R.T. Qiu, D.X. Fan, A. Liu, Exploring the Booking Determinants of the Airbnb Properties: An Example of the Listings of London, in: B. Stangl, J. Pesonen (Eds.), Information and Communication Technologies in Tourism 2018, Springer, 2018, pp. 44–51.
[37] J. Lyu, M. Li, R. Law, Experiencing P2P accommodations: Anecdotes from Chinese customers, Int. J. Hosp. Manag. 77 (2019) 323–332.
[38] L. Kwok, K.L. Xie, Pricing strategies on Airbnb: Are multi-unit hosts revenue pros? Int. J. Hosp. Manag. 82 (2019) 252–259.
[39] L. Masiero, J.L. Nicolau, R. Law, A demand-driven analysis of tourist accommodation price: A quantile regression of room bookings, Int. J. Hosp. Manag. 50 (2015) 1–8.
[40] H. Öğüt, B.K. Onur Taş, The influence of internet customer reviews on the online sales and prices in hotel industry, Serv. Ind. J. 32 (2) (2012) 197–214.
[41] E.N. Torres, D. Singh, A. Robertson-Ring, Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry, Int. J. Hosp. Manag. 50 (2015) 77–83.
[42] D. Gavilan, M. Avello, G. Martinez-Navarro, The influence of online ratings and reviews on hotel booking consideration, Tour. Manag. 66 (2018) 53–61.
[43] C.-C. Chen, Y.-C. Chang, What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness, Telematics Inform. 35 (5) (2018) 1512–1523.
[44] Y. Ju, K.-J. Back, Y. Choi, J.-S. Lee, Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction, Int. J. Hosp. Manag. 77 (2019) 342–352.
[45] M. Cheng, X. Jin, What do Airbnb users care about? An analysis of online review comments, Int. J. Hosp. Manag. 76 (2019) 58–70.
[46] I. Altman, D.A. Taylor, Social penetration: The development of interpersonal relationships, Rinehart & Winston, Holt, 1973.
[47] J.L. Gibbs, N.B. Ellison, R.D. Heino, Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in internet dating, Commun. Res. 33 (2) (2006) 152–177.
[48] H. Moon, L. Miao, L. Hanks, N.D. Line, Peer-to-peer interactions: Perspectives of Airbnb guests and hosts, Int. J. Hosp. Manag. 77 (2019) 405–414.
[49] I.P. Tussyadiah, Factors of satisfaction and intention to use peer-to-peer accommodation, Int. J. Hosp. Manag. 55 (2016) 70–80.
[50] I.P. Tussyadiah, An exploratory study on drivers and deterrents of collaborative consumption in travel, Information and Communication Technologies in Tourism 2015, Springer, 2015, pp. 817–830.
[51] B. Cova, D. Dalli, D. Zwick, Critical perspectives on consumers' role as 'producers': Broadening the debate on value co-creation in marketing processes, Mark. Theory 11 (3) (2011) 231–241.
[52] I.P. Tussyadiah, J. Pesonen, Impacts of peer-to-peer accommodation use on travel patterns, J. Travel Res. 55 (8) (2016) 1022–1040.
[53] A.-G. Johnson, B. Neuhofer, Airbnb–an exploration of value co-creation experiences in Jamaica, Int. J. Contemp. Hosp. Manag. 29 (9) (2017) 2361–2376.
[54] J.W. Pennebaker, R.L. Boyd, K. Jordan, K. Blackburn, The Development and Psychometric Properties of LIWC2015, University of Texas at Austin, Austin, TX, 2015.
[55] A.R. Walls, A cross-sectional examination of hotel consumer experience and relative effects on consumer values, Int. J. Hosp. Manag. 32 (2013) 179–192.
[56] K.L. Xie, L. Kwok, The effects of Airbnb's price positioning on hotel performance, Int. J. Hosp. Manag. 67 (2017) 174–184.
[57] C.B. Dean, Testing for Overdispersion in Poisson and Binomial Regression Models, J. Am. Stat. Assoc. 87 (418) (1992) 451–457.
[58] R. Winkelmann, Econometric Analysis of Count Data, Springer Science & Business Media, 2013.
[59] A.C. Cameron, P.K. Trivedi, Econometric models based on count data. Comparisons and applications of some estimators and tests, J. Appl. Econ. 1 (1) (1986) 29–53.
[60] B. Monty, M. Skidmore, Hedonic Pricing and Willingness to Pay for Bed and Breakfast Amenities in Southeast Wisconsin, J. Travel Res. 42 (2) (2003) 195–199.
[61] R.B. Ekelund Jr., R.F. Hébert, A history of economic theory and method, Waveland Press, 2013.
[62] D.J. Egan, K. Nield, Towards a theory of intraurban hotel location, Urban Stud. 37 (3) (2000) 611–621.
[63] A. Marshall, Principles of economics: unabridged eighth edition, Cosimo, Inc.,

2009.

[64] G.A. Akerlof, The market for “lemons”: Quality uncertainty and the market mechanism, in: P Diamond, M Rothschild (Eds.), Uncertainty in Economics, Elsevier, 1978, pp. 235–251.

[65] V. Kakar, J. Voelz, J. Wu, J. Franco, The Visible Host: Does race guide Airbnb rental rates in San Francisco? J. Hous. Econ. 40 (2018) 25–40.

[66] N. Su, A.S. Mattila, Does gender bias exist? The impact of gender congruity on consumer’s Airbnb booking intention and the mediating role of trust, Int. J. Hosp. Manag. 102405 (2019), https://doi.org/10.1016/j.ijhm.2019.102405.

Baidyanath Biswas is an Assistant Professor of Information Technology at the International Management Institute Kolkata. He received his Ph.D. from the Indian Institute of Management Lucknow (IIM Lucknow) in 2019. Before joining IIM Lucknow in 2014, he worked for nine years with Infosys Technologies, IBM India, and Cognizant Technology Solutions as a Mainframe and DB2 programmer. His research has been published in the proceedings of top peer-reviewed international conferences, namely, AMCIS (2016 until 2019), HICSS (2018), and Pre-ICIS Workshops (2018). He has also published in *Computers in Industrial Engineering* , and the *Journal of Enterprise Information Management* . He serves as a reviewer for *Electronic Markets*, *Information Systems Frontiers*, *ACM SIGMIS Database*, *International Journal of Production Economics*, *Computers in Industrial Engineering*, and *Journal of the Operational Research Society*. Currently, he serves as the Associate Editor of the *Global Business Review* (SAGE) journal.

Pooja Sengupta is a researcher with interests in inflated discrete distributions, multivariate discrete distributions, generalized linear models, longitudinal data analysis. She has over 6 years of experience in teaching courses in Statistics such as linear algebra, linear model, regression analysis, statistical computation using R and Data Analytics using R. She has published in *Communications in Statistics-Theory and Methods*, *Journal of Statistical Theory and Practice,* and the *Journal of Statistical Computation and Simulation.* She has also presented research papers at different prestigious international and national conferences; the most recent one is the 62nd World Statistics Congress, organized by International Statistical Institute in Kuala Lumpur. Apart from this, she has also presented a research paper at the International Indian Statistical Association, 2014 Conference at Riverside, California, and 2014 Conference of Virginia Academy of Science Chapter, American Statistical Association. She is also involved in a UGC project funded by the Ministry of Human Resource Development, where she is the content writer of two papers on regression models in the Statistics post-graduate curriculum. She has received her bachelor’s degree, B.Sc. Statistics (Honours) and M.Sc. in Statistics from the University of Calcutta and joined the Old Dominion University, Virginia, the USA in 2010 from where she received her Ph.D. in Computational and Applied Mathematics with a concentration in Statistics.

Dwaipayan Chatterjee is a student at Department of Statistics, University of Calcutta pursuing his Master's degree. His areas of interest are Regression, Time Series, Neural Networks and Measure Theory.