

What Drives Airbnb Customers' Satisfaction in Amsterdam? A Sentiment Analysis

Heyam Abdullah Bin Madhi¹, Muna M. Alhammah²

Management Information System Dept
King Saud University, Riyadh
Saudi Arabia

Abstract—The sharing economy is a new socio-economic system that allows individuals to rent out their personal belongings, such as their private car or a room in their home, for a short period. This study aims to investigate the attributes that impact customers' satisfaction when using the sharing economy propriety rentals websites. Large data sets of Airbnb's online reviews and listings in Amsterdam were analyzed using sentiment analysis, word clustering, ordinal logistic regression, and visualization techniques. Findings reveal that the polarity of Airbnb guests reviews in Amsterdam is significantly impacted by property price, value, cleanliness, rate, host communication, easiness of check-in, the accuracy of property description, and whether the host is super host or not. Surprisingly, the property neighborhood was not found to impact customers' sentiment in Amsterdam. In addition, Airbnb guests in Amsterdam tend to positively express their experience satisfaction level mainly based on property exact location and host interaction followed with the facilities surrounding the property, property cleanliness, and room quality. On the other hand, negative online reviews tend to be mainly linked to problems with check-in services followed by aspects related to weak host interaction, location, and room quality. The results indicate that Airbnb hosts need to offer clear and easy check-in services with emphasizing the importance of keeping a good communication channel with their guests to enhance customers' experience and increase customers' satisfaction level. Future studies should investigate the applicability of the findings of this study in the context of other cities.

Keywords—Airbnb; customer satisfaction; customer experience; big data; sentiment analysis; ordinal logistic regression

I. INTRODUCTION

The development of media technology has contributed to the prosperity of a new business concept, which is the sharing economy [1]. A sharing economy is an advanced approach for conducting businesses, where people can share their own resources in a decentralized or peer-to-peer (P2P) platform for an exchange of money [2]. It has generated new business opportunities, especially for those who don't have the essential resources. Additionally, property owners have granted the chance to liquidate and utilize their unused assets across the globe, allowing them to serve customers from different regions and cultures [3]. The sharing economy had provided the necessary services in different qualities and prices 24/7 in order to provide suitable services to fulfill individual needs and cater to a different level of purchasing power. As a result, consumption and production sustainability have been accelerated [4]. The concept of the sharing economy has been

applied in different fields including, shared mobility, hospitality, on-demand staffing, and media streaming. Lime, JustPark, and Uber are some examples of sharing economy applications. In the context of hospitality, several companies have emerged, such as Airbnb, Couchsurfing, HomeAway, Roomorama, and HomeExchange. These organizations act as an intermediary between guests and hosts to facilitate communications between them [5].

Looking at Airbnb as an alternative option for hotels, several existing research have explored this platform from different perspectives. The author in [6] examined the substitution and complementary effects of Airbnb supply on hotel sales performance patterns while [7] studied the extent to which Airbnb is used as a substitute for hotels and [8] studied the price difference between hotel property and the nearby Airbnb listings offers. Additionally, many other researchers have studied the social contact of guests in the Airbnb accommodation. For instance, [9] explored the social contact of Airbnb guests during their stay, taking into account these types of contact: guest to host, guest to the community, and guest to guest communications whereas [10] studied the reciprocal aspect of social interactions in P2P accommodations. Several other studies have examined factors that might affect guests' booking intentions such as studying the impact of gender congruity between guests and hosts on guests' booking [11], and the impact of properties and hosts descriptions on guests' booking and posting reviews [12]. Moreover, some authors recognized the importance of studying online advertising strategies [13], price determination [14], trust-attachment building mechanisms [15], and customers' psychological behavior [16]. Despite the diversity of Airbnb research fields, few studies have examined guests' reviews toward their Airbnb stay. [17] have investigated the attributes that influence Airbnb users' experiences through the use of text mining and sentiment analysis. Similarly, [18] have investigated hidden dimensions in textual reviews through the use of Latent Aspect Rating Analysis (LARA). Analyzing big data such as customer reviews is a key success factor for sharing economy businesses [19]. It enables a business owner to understand the organization's strengths and weaknesses along with customers' habits. Therefore, it helps to predict future trends, improve decision making, detect frauds, and eventually increase business revenues.

However, to the best of the authors' knowledge, there are no previous studies investigating factors that drive customers' satisfaction through analyzing guests' online reviews

simultaneously with the listing features. Therefore, to better understand customers' needs and achieve customers' satisfaction, there is a need to identify factors that makes customers satisfied. Analyzing users' comments and reviews, through the use of various data mining techniques, helps to identify factors impacting users' satisfaction [19]. Therefore, the research question to be answered in this research is:

What makes customers satisfied when using the Airbnb platform in Amsterdam?

Airbnb online reviews dataset in Amsterdam will be analyzed simultaneously with the listings details dataset by using sentiment analysis, topic clustering, regression, and visualization techniques. The paper begins with reviewing relevant literature, clarifying the applied methodology, presenting the results and discussion, and finally concludes the paper.

II. LITERATURE REVIEW

A. Sharing Economy

According to [20], sharing economy concept is the result of a massive shift, not a new phenomenon in itself. This is because sharing economy is an online community that includes economic aspects such as selling, buying, and renting which already existed before. However, the new concept is distinguished by allowing users to participate using different types of resources, in both economic and social aspects, whether it was a human, merchandise, service, or property [21]. A more comprehensive description was provided by [22], where they defined sharing economy as a P2P environment that enables a person to be a client and a provider at the same time. The sharing economy facilitates social participation and cooperative consumption which in turn supports protecting the environment, minimizing resource waste, and enhancing community awareness [4]. The sharing economy covers five main areas: product-service systems, redistribution markets, collaborative lifestyles, access-based consumption, and commercial sharing systems.

As mentioned earlier, sharing economy businesses have witnessed rapid development in the last decade. The author in [5] justify the reason to be related to the fact that sharing economy businesses are less expensive for consumers, especially in the case of accommodation as it is cheaper for travelers to choose a place from Airbnb to stay, rather than renting a room in a hotel. Additionally, [19] have demonstrated that science and technology revolutions are playing a huge role in sharing economy growth through the development of electronic social networking (e-SN) platforms which contributed to the development of society and the economy.

Although business fields differ, online hospitality is one of the most famous fields in sharing economy. The word hospitality was defined by [22] as two associated meanings. From the guests' perspective, hospitality is a process of delivering a service of high quality while from hosts' perspectives; it is about providing services and rooms with a focus on profitability. In terms of hospitality, using the sharing economy, hospitality means allowing individuals to invest their properties to host visitors for a small financial return through

the use of online sharing economy portals such as the Airbnb website.

B. Customers' Satisfaction

The word satisfaction is defined as "a psychological statement with the fulfillment of a need or desire and the pleasure obtained by such fulfillment" [19]. Customers' satisfaction can be interpreted as customers' judgments, opinions, feelings, impressions, or emotional reactions towards their comprehensive experiences of a product or a service [23]. [22] have indicated that it is essential to understand customers' satisfaction and how it is impacted to ensure customers' commitments and obtaining their loyalties. Increasing customer satisfaction will increase customer retention and loyalty. Consequently, organizations' revenues and profits will rise as there is a direct link between customers' satisfaction and profit [24]. However, increasing customers' satisfaction comes at a cost, so organizations must be careful not to reach excessive satisfaction that costs them a lot without returns. To achieve the maximum customer satisfaction level with minimum cost, organizations must understand their customers and what makes them satisfied [24].

Obtaining customers' satisfaction could be a big challenge for many organizations. The first step to increase customers' satisfaction is identifying customers' current satisfaction levels and the reasons that led to this satisfaction level. Capturing and measuring customer experience is an excellent way to achieve that. Analyzing rating scores and comments that were produced by customers after the end of a service, can grant comprehensive knowledge about customers' satisfaction [25]. In this research, Airbnb guests' reviews will be analyzed, to better understand customers' satisfaction along with factors that affect their level of satisfaction.

C. Airbnb

Airbnb, Inc. is an online marketplace that was founded in 2008 for offering accommodates, homestays, or tourism experiences. It is one of the most leading and profitable platforms within the sharing economy [26]. The organization works as a broker that connects real estate owners with hospitality seekers. In particular, it links individuals who need to lease out their homes (hosts) with individuals who are searching for lodging (guests) in that region in exchange for a small commission from each reservation [27]. Hosts are taking part in Airbnb as an approach to obtain some pay from their property even though they are facing the risk that guests may harm their property. On the other hand, guests are taking advantage of having accommodation with relatively lower prices than other places while at the same time facing the risk that the property's quality will not be as expected. Generally, Airbnb provides services for both guests and hosts to provide a better coordinating outcome [28].

Several types of research have been applied to Airbnb, some focusing on user experience and multimedia, others on administration, tourism, and architecture [29]. Covering multiple areas including review bias, hospitality exchanges, price and neighborhood prediction, neighborhoods ranking, socio-economic characteristics, listing recommendation system, rentals' distribution, users' preferences and expectations, image mining, demand mining, grading schema,

matching schema, consumer segmentation, trust evaluation, innovation adoption, and adoption evaluation. In addition to some researches that focus on analyzing customers' satisfaction dimensions [30], analyzing customers feedback using text mining [3], and analyzing public opinions using content analysis [21]. Previous studies have almost exclusively focused on analyzing user reviews separately without linking them to listings data. A more comprehensive analysis is required for analyzing factors affecting customers' satisfaction, in a way that enables non-technical users to understand it without any effort.

III. METHODOLOGY

A. Data Collection and Preparation

Over time, extensive literature has confirmed the importance and usefulness of analyzing online review comments for researchers, business owners, as well as other customers. Since users' comments illustrate their perceptions and feelings formed while using a service or a product, it could be used to measure and analyze users' satisfaction level, along with identifying key factors that led to such a level of satisfaction [31].

On the Airbnb website, guests are posting reviews that reflect their experiences and opinions about the property they stay on. This provides data that can be analyzed, to identify factors that make customers satisfied [32]. The current study obtained online review comments of Airbnb' guests about their accommodations in Amsterdam, as well as detailed listing data from the Inside Airbnb website [33]. Given the fact of the massive availability of Airbnb data as public Datasets available on the Inside Airbnb website, both datasets were obtained easily. The listing dataset contains 107 attributes that describe host, property features, and scored reviews while the review dataset contains six attributes and more than 400 thousand review records. Datasets were then pre-processed or cleaned before the analysis phase, as performing this step usually helps to generate feasible results [34]. For the listings dataset, the cleaning process was done by eliminating attributes containing errors, no values (blank), and useless values. Thereafter, new columns were generated or derived from existing attributes, such as host experience in months derived from the host starting date, weekly and monthly discounts derived from prices. On the other hand, the review dataset was checked for missing data and non-English content to be removed. Google Online Spreadsheet was used to detect comments' language, and to clean non-English data from the set, which represents 18% of the records. Unwanted columns were eliminated including comment ID, reviewer ID, and reviewer name. Additionally, a filter was applied using the date of the comment, choosing comments within 2019 only, to minimize the number of reviews and make it manageable to analyze the data using home devices with limited capabilities. Dummy comments containing characters or numbers only were also removed leaving a total of 110,747 comments eligible for the analysis.

B. Data Analysis

1) *Sentiments analysis*: The first step in the analysis stage began with sentiment analysis. The sentiment analysis has

been conducted using MeaningCloud's Text Analytics add-in in Excel. MeaningCloud sentiment analysis produces two sheets. The first one is for the global sentiment analysis sorting includes the polarity, i.e., positive (P), negative (N), or Neutral (NEW), and a confidence score while the other one is for comments' topics sorting including topics categories and topics types (which were used later on in text mining).

2) *Regression analysis*: The second step in the analysis was identifying factors affecting users' satisfaction. To begin this step, the two datasets were merged. Ordinal logistic regression (OLR) test was chosen as suggested by [35], to answer the research question, by determining which of the independent variables (IV) have a statistically significant effect on our dependent variable (DV), which is polarity.

The reason for selecting the OLR test in this study is because the DV is an ordinal categorical variable where the value is represented with three points scale (N=0, NEW=1, and P=2) and the IVs are a mix of continuous and categorical. Multicollinearity test was conducted to ensure there are no highly correlated IVs, and only two IVs (accommodates and beds) were highly correlated with a result of >0.8; accordingly, the beds variable was removed from the test. Without performing the multicollinearity test, there would be a problem with determining the variable that contributes to the interpretation of the DV [35]. After eliminating the highly correlated variables, the final variables set contained 32 attributes as following: host experience in months, host response time, host response rate, the host is super-host, host total listings count, the host has profile pic, host identity verified, neighborhoods cleansed, is location exact, property type, room type, accommodates, bathrooms, bedrooms, bed type, square feet, price, weakly discount, monthly discount, security deposit, cleaning fee, extra people fee, instant bookable, cancellation policy, require guest phone verification, review scores rating, review scores accuracy, review scores cleanliness, review scores check-in, review scores communication, review scores location, and review scores value.

Taking the results from the OLR test, factors that affect polarity were visualized using Power BI software. Power BI was also used to conduct a text mining test to find the most frequent words within comments. Word's visualization helped to derive additional insights about repeated words in different polarity levels.

IV. FINDINGS AND DISCUSSION

A. Sentiment Analysis and Topic Clustering

The results of the sentiment analysis indicate that Airbnb users were mostly having a positive experience in Amsterdam (see Fig. 1). The topic clustering result shows that users were overwhelmingly positive about two aspects of their experience i.e. location and host (Table I). For instance, the likelihood score of 52% for a location means that reviews with the term location represent 52% of the positive sentiments. Location is an essential indicator for customers' satisfaction since they care about the proximity of different facilities to their residence such as public transportation, restaurants, park, station,

markets, and shops. As a second major topic within the positive sentiments, the term host represents 31.66% of the positive results; this indicates that many users were satisfied with the way their hosts were treating them. In contrast, the topic that consistently received a negative sentiment was check-in (Table II). The term Check-in represents the majority of the negative sentiment with a likelihood score of 95.69%. This result gives an indicator for hosts to improve their check-in services, to make this process quick, efficient, and smooth. This can be achieved by following the authoritative guide to the Airbnb check-in process. Fig. 2 shows word clouds that support the previous findings.

Although Airbnb's comments are mainly positive, taking a closer look at areas where negative sentiments have occurred, allows Airbnb hosts to address these areas by fixing any problems or setting future expectations (e.g., improving check-in services). Moreover, the strong appearance of topics related to the city's environment (such as location, facilities, transportation, restaurants, parks, stations, markets, and shops) among positive sentiments compared to their weak presence among negative ones, indicating that Amsterdam's general environment can play an essential role in shaping the Airbnb guests positive experience.

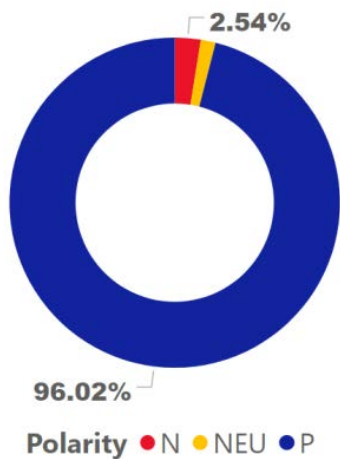


Fig. 1. Polarity Likelihood, Where N is Negative, NEU is Neutral, and P is Positive.



Fig. 2. Word Cloud for Airbnb Positive (Left) and Negative Reviews (Right).

TABLE I. WORDS' LIKELIHOOD SCORES OF SENTIMENT ANALYSIS FOR AIRBNB USERS' ONLINE REVIEWS

Positive			Negative		
Words	Likelihood	Count	Words	Likelihood	Count
Location	52.47%	52554	Check-in	95.69%	41473
Host	31.66%	31708	Host	1.63%	707
Facility	3.51%	3518	Location	0.84%	366
Clean	1.70%	1704	Room	0.35%	151
Room	1.50%	1504	Facility	0.30%	131
Walk	1.23%	1233	Bed	0.19%	81
Transportation	1.21%	1208	Walk	0.13%	58
Bed	1.18%	1185	Transportation	0.12%	51
Great	0.77%	767	Other Entity	0.09%	39
Top	0.57%	571	Service	0.06%	26
Restaurants	0.39%	395	Kitchen	0.06%	25
Process	0.32%	321	Top	0.06%	25
View	0.28%	282	Restaurants	0.05%	23
Other Entity	0.27%	274	Noise	0.05%	22
Park	0.27%	270	Station	0.05%	21
Amenities	0.26%	265	Amenities	0.04%	16
Station	0.26%	262	View	0.04%	16
Ship	0.24%	244	Water Form	0.03%	15
Service	0.20%	198	Shower	0.03%	14
Family	0.19%	190	Not-Clean	0.03%	13
Water Form	0.17%	175	Style	0.02%	10
Couple	0.16%	160	Couple	0.02%	8
Market	0.14%	143	Animal	0.02%	7
Style	0.14%	143	Park	0.02%	7
Check-in	0.14%	142	Market	0.01%	6
Kitchen	0.13%	132	Process	0.01%	5
Breakfast	0.12%	118	Coffee-Shop	0.01%	4
Coffee-Shop	0.10%	104	Family	0.01%	4
Shower	0.10%	99	Ship	0.01%	4
Animal	0.06%	56	Breakfast	0.01%	3
Quiet	0.05%	53	Shops	0.01%	3
House	0.03%	35	Airport	0.00%	2
Shops	0.03%	33	Event	0.00%	2
Staff	0.03%	32	Open	0.00%	1
Airport	0.02%	18	Plant	0.00%	1
Museum	0.01%	15	Sports	0.00%	1
Plant	0.01%	12			
Sports	0.01%	10			
Activities	0.01%	9			
Event	0.01%	7			
Open	0.00%	3			

TABLE II. ORDINAL REGRESSION RESULTS: PARAMETER ESTIMATES

Description	Attribute	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Sentiment threshold where 0 is neutral	[Polarity = 0]	-16.742	9398.157	.000	1	.999	-18436.79	18403.307
The length of the time since a host registered on Airbnb (#months).	Host_Experience_in Months	-.030	.034	.764	1	.382	-.096	.037
The rate of one host responding to his guests	host_response_rate	.070	.060	1.343	1	.247	-.048	.188
The number of one host total listings	host_total_listings_count	.083	.149	.311	1	.577	-.209	.376
The number of guests accommodated by a given listing.	Accommodates	-.568	.468	1.471	1	.225	-1.485	.350
Number of bathrooms	Bathrooms	-.307	1.012	.092	1	.761	-2.291	1.676
Number of bedrooms	Bedrooms	.705	.683	1.065	1	.302	-.634	2.043
Listing price in dollar	Price	-.018	.007	5.957	1	.015*	-.032	-.003
Discount rate per week	Discount_Rate	.004	.039	.011	1	.918	-.073	.081
Discount rate per month	Discount_Monthly	.003	.029	.014	1	.905	-.054	.061
Some hosts require a security deposit for their listing	security_deposit	.003	.002	1.655	1	.198	-.002	.008
Extra fee for cleaning services	cleaning_fee	.015	.014	1.271	1	.260	-.011	.042
Extra fee for having extra people to reside in the property	extra_people_fee	.004	.013	.079	1	.779	-.022	.030
Property size in f ²	square_feet	1.614E-5	.001	.001	1	.982	-.001	.001
Within an hour = 1, a few hours = 2, hours = 3, a few days = 4	[host_response_time=1]	4.585	5.982	.587	1	.443	-7.140	16.310
	[host_response_time=2]	1.117	.987	1.279	1	.258	-.819	3.052
	[host_response_time=3]	.575	.766	.563	1	.453	-.926	2.076
	[host_response_time=4]	0 ^a	.	.	0	.	.	.
Each code represents a specific neighbourhood area within Amsterdam.	[neighbourhood_cleansed=1]	11.325	3394.217	.000	1	.997	-6641.218	6663.868
	[neighbourhood_cleansed=3]	.967	1.996	.235	1	.628	-2.946	4.880
	[neighbourhood_cleansed=4]	11.666	6271.111	.000	1	.999	-12279.48	12302.817
	[neighbourhood_cleansed=5]	-1.291	1.681	.589	1	.443	-4.586	2.005
	[neighbourhood_cleansed=6]	.726	1.731	.176	1	.675	-2.667	4.120
	[neighbourhood_cleansed=8]	.974	1.731	.316	1	.574	-2.419	4.366
	[neighbourhood_cleansed=9]	.017	1.620	.000	1	.992	-3.159	3.193
	[neighbourhood_cleansed=10]	-3.616	2625.809	.000	1	.999	-5150.107	5142.876
	[neighbourhood_cleansed=11]	17.599	2158.338	.000	1	.993	-4212.666	4247.863
	[neighbourhood_cleansed=12]	1.179	.000	.	1	.	1.179	1.179
	[neighbourhood_cleansed=13]	-.669	2.339	.082	1	.775	-5.254	3.915
	[neighbourhood_cleansed=14]	.286	2.058	.019	1	.889	-3.748	4.320
	[neighbourhood_cleansed=15]	.466	1.943	.057	1	.811	-3.343	4.275
	[neighbourhood_cleansed=16]	13.494	2137.889	.000	1	.995	-4176.692	4203.680
[neighbourhood_cleansed=17]	-1.021	1.821	.314	1	.575	-4.590	2.549	
[neighbourhood_cleansed=18]	0 ^a	.	.	0	.	.	.	
1= Private room in	[property_type=1]	-14.115	9398.155	.000	1	.999	-18434.16	18405.930

Description	Attribute	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
apartment, 2= Private room in townhouse, 3= Private room in houseboat, 4= Entire guest suite, 5= Boat, 6= Private room in bed and breakfast, 7= Private room in guesthouse, 8= Houseboat, 9= Entire villa, 10= Entire apartment, 11= Entire townhouse, 12= Entire loft, 13= Entire serviced apartment, 14= Entire house	[property_type=2]	-12.641	9398.155	.000	1	.999	-18432.68	18407.405
	[property_type=3]	-14.589	9398.155	.000	1	.999	-18434.63	18405.456
	[property_type=4]	0 ^a	.	.	0	.	.	.
	[property_type=5]	-15.302	9398.155	.000	1	.999	-18435.34	18404.744
	[property_type=6]	-15.791	9398.155	.000	1	.999	-18435.83	18404.255
	[property_type=7]	3.182	9437.052	.000	1	1.00	-18493.10	18499.464
	[property_type=8]	-14.438	9398.155	.000	1	.999	-18434.48	18405.607
	[property_type=9]	2.391	9522.100	.000	1	1.00	-18660.58	18665.365
	[property_type=10]	3.605	11997.929	.000	1	1.00	-23511.90	23519.113
	[property_type=11]	-.888	.000	.	1	.	-.888	-.888
	[property_type=12]	-14.108	9398.155	.000	1	.999	-18434.15	18405.938
	[property_type=13]	-2.049E-6	10602.550	.000	1	1.00	-20780.616	20780.616
	[property_type=14]	0 ^a	.	.	0	.	.	.
	Shared room = 1, private room = 2 and entire home = 3.	[room_type=1]	2.022	1.056	3.667	1	.056	-.048
[room_type=2]		-1.423	1.668	.727	1	.394	-4.692	1.847
[room_type=3]		0 ^a	.	.	0	.	.	.
Airbed=1, Real bed=2, Couch=3	[bed_type=1]	.842	3.396	.061	1	.804	-5.813	7.497
	[bed_type=2]	14.717	3255.619	.000	1	.996	-6366.178	6395.612
	[bed_type=3]	0 ^a	.	.	0	.	.	.
Cancellation policies: flexible = 0, moderate = 1 and strict = 2.	[cancellation_policy=0]	-.474	1.543	.094	1	.759	-3.498	2.551
	[cancellation_policy=1]	-.625	1.450	.186	1	.666	-3.468	2.217
	[cancellation_policy=2]	0 ^a	.	.	0	.	.	.
A host with the superhost badge is labelled 0 and 1 otherwise.	[host_is_superhost=0]	-2.425	.815	8.848	1	.003**	-4.023	-.827
	[host_is_superhost=1]	0 ^a	.	.	0	.	.	.
Host picture presented	[host_has_profile_pic=1]	0 ^a	.	.	0	.	.	.
Verified host = 1, non-verified host = 0	[host_identity_verified=0]	1.076	.862	1.557	1	.212	-.614	2.766
	[host_identity_verified=1]	0 ^a	.	.	0	.	.	.
Exact location=0, not exact location=1	[is_location_exact=0]	1.952	1.919	1.034	1	.309	-1.810	5.714
	[is_location_exact=1]	0 ^a	.	.	0	.	.	.
Instant bookable=0, not instant bookable=1	[instant_bookable_A=0]	-.590	.732	.648	1	.421	-2.025	.846
	[instant_bookable_A=1]	0 ^a	.	.	0	.	.	.
Require guest phone verification=0, not required guest phone verification=1	[require_guest_phone_verification_A=0]	.315	.970	.106	1	.745	-1.585	2.216
	[require_guest_phone_verification_A=1]	0 ^a	.	.	0	.	.	.
Review ranges from 1 to 5 where 1 is the lowest rate and 5 is the highest rate	review_scores_rating	.035	.006	38.512	1	.000***		
	review_scores_accuracy	.168	.040	17.839	1	.000***		
	review_scores_cleanliness	.280	.029	94.662	1	.000***		
	review_scores_checkin	.142	.039	13.321	1	.000***		
	review_scores_communication	.176	.039	20.379	1	.000***		
	review_scores_location	.028	.030	.886	1	.347		
	review_scores_value	.095	.035	7.458	1	.006**		

B. Factors Influencing Comments Polarity

After applying an OLR test, the authors found that only nine out of thirty-two factors were influencing customers' satisfaction with a significant P value of < 0.05 (Table II). These affecting factors are price, room type (Entire Home/Apartment), the host is not a super-host, review scores rating, review scores accuracy, review scores cleanliness, review scores check-in, review scores communication, and review scores value. Surprisingly, the property neighborhood was not found impacting customers sentiment in Amsterdam which might be a case only limited to Amsterdam as it is one of the safest cities in the world.

1) *Price:* Price is the first-factor influences customers' satisfaction (B= 5.957, p<0.05). Fig. 3 illustrates the relationship between properties' daily prices and the polarity of comments sentiment. The figure shows that as the price raises, the negative reviews decline. In other words, positive comments are more commonly associated with high price properties comparing with negative ones. According to [14], properties with some features such as luxurious, penthouse, unique, chic & designed, duplex, Sauna, Jacuzzi & Spa, are relatively more expensive than other properties. It is, hence, expected that the benefits of obtaining luxurious characteristics, make customers willing to pay more.

Similarly, [36] indicates that some amenities are considered a price determinant. The availability of these amenities will give Airbnb customers the quality they are seeking. Eventually, they will be more satisfied and happier compared to customers who didn't obtain these amenities. Therefore, property price impacts customers' satisfaction indirectly since prices are determined by quality, which is critical to gain customers' satisfaction.

2) *Room type:* The second affecting factor is the room type, including the entire home/apartment, hotel room, private room, and shared room. Fig. 4 showing the percentage of negative and positive comments among the four property types (calculated as a percentage out of the same polarity). For example, entire home/apt negative comments represent 29% out of the total negative sentiments for the four types. The type of entire home/apt is more likely to get negative reviews than positive ones. After comparing the average of different attributes based on user sentiment (Positive versus Negative), the researchers found some factors that affect the entire home/apt polarity (see Fig. 5). The result indicates that decreasing the average square feet influences making negative feelings. Moreover, as host listing increased, the probability of getting a dissatisfied guest will be increased as well. Hence, it is evidence that there's a negative relationship between host total listings and customers' satisfaction. This could be due to hosts being busy handling all their properties simultaneously, therefore providing poor customer services. Eventually, hosts will fail to deliver the expected level of care to their guests.

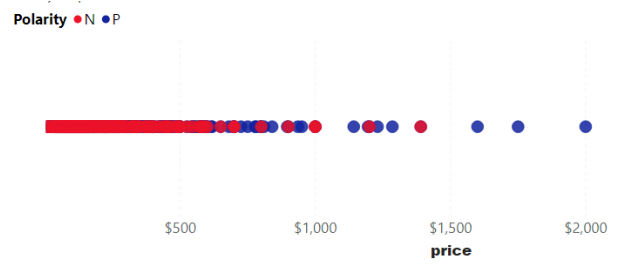


Fig. 3. Price Distribution for Negative and Positive Polarity.

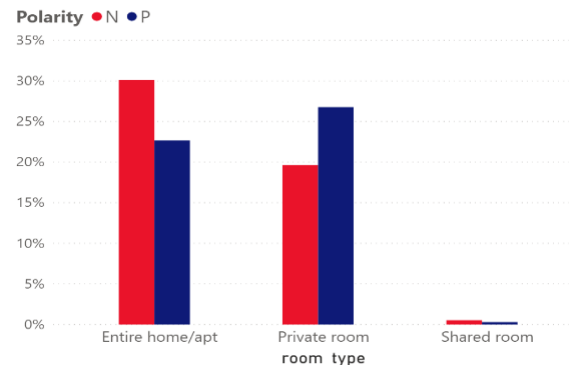


Fig. 4. Likelihood of Polarity, Grouped by Room Type.

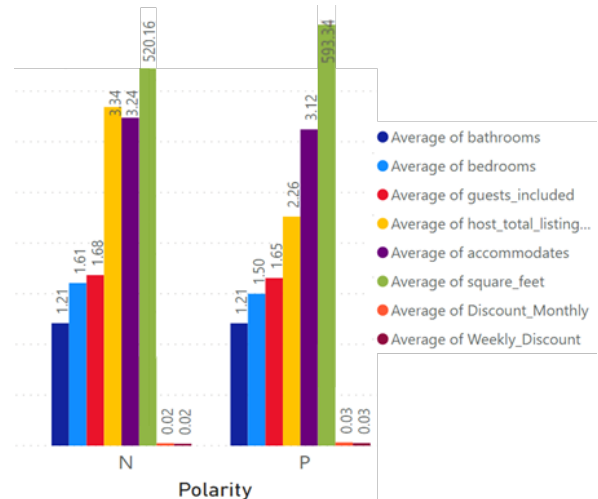


Fig. 5. Factors Influencing users' Polarity when Renting Entire Home/Apt.

3) *Super-host:* The third affecting factor is when the host is not a super-host. Fig. 6 is showing a similar distribution of super-host within the positive comments, which illustrates that whether a host is super-host or not, it will not affect the positivity of a sentiment. In contrast, the negative sentiments are widely affected by the host being not a super-host. Fig. 6 illustrated that 88.8% of unsatisfied customers were hosted by non-super-hosts. This supports [24] findings, who confirmed that poor services have a more significant impact on satisfaction level than good services. People tend to expect a certain level of quality. Obtaining this level will not positively

affect their satisfaction degree, but failing to reach this level will significantly affect their satisfaction negatively. According to [10], what distinguishes P2P accommodation from traditional hotels is to create a sense of place by providing a closer interaction with a resident (host). In such an aspect, online and face-to-face interactions with the host can influence customers' satisfaction, due to the atmosphere host creates. This fact supports our contribution here since low host interactions will lead to customer dissatisfaction.

4) *Neighborhood*: Surprisingly, though location plays an important role as an affecting element on polarity within topic clustering results (see Fig. 2), the OLR test provided evidence that neighborhood wasn't affecting the polarity (Table II). After a careful review of the results, it seems that the neighborhood by itself has no impact, while the exact location inside each neighborhood does. The reason for this situation is that different location within the same neighborhood district in terms of facilities, quietness, transportation, etc. To confirm OLR result, the relationship between polarity and neighborhoods were represented in Fig. 7 using maps and a bar chart. The maps show a similar distribution between the Ps and Ns comments, and the bar chart supported this result by indicating that each neighborhood has almost the same distribution rate of the Ps and Ns (top 15 neighborhoods). This could be due to the characteristic that made Amsterdam a unique city that everyone would like to visit. Such as its high quality of life, the European culture lifestyle, the small village feeling, and its beautiful landscape in all neighborhoods. One of these characteristics is safety, as [37] illustrated that Amsterdam is considered the fourth safest city in the world and the safest among all of Europe. The level of safety of all Amsterdam's neighborhoods is high in terms of personal security, health security, infrastructure security, and digital security. This support our findings that Amsterdam's neighborhoods do not affect Airbnb customer's satisfaction since the visitors (guests) will be positively biased toward this city and all of its neighborhood in general.

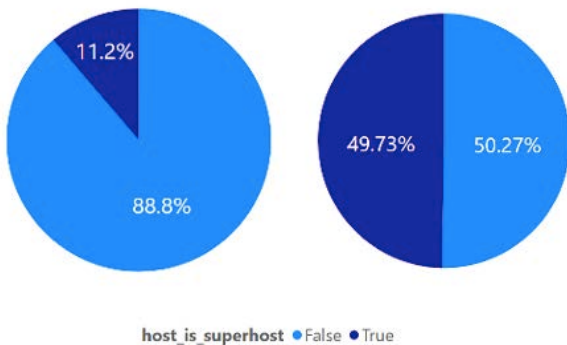


Fig. 6. Superhost Proportion in Positive Comments (Right) and in Negative Comments (Left).

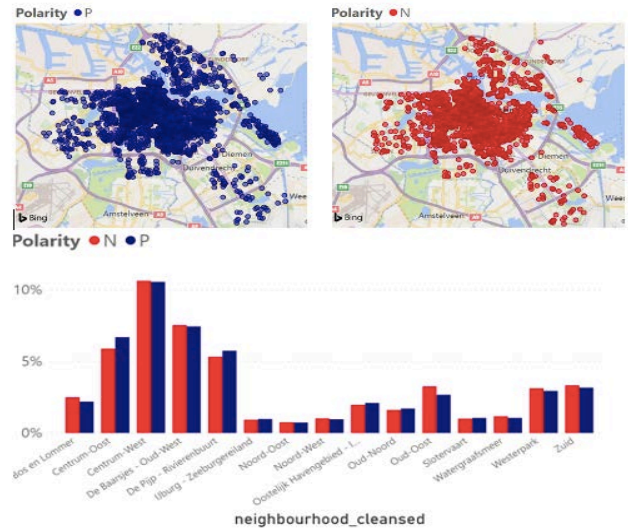


Fig. 7. Polarity Distribution based on Locations.

V. CONCLUSION AND FUTURE WORK

Airbnb represents a flourishing implementation of the sharing economy generally, and hospitality mainly. While Airbnb remains a topic of significant attention within the sector, preliminary literature on Airbnb has examined customers' reviews independently. This study is the first endeavor to analyze Airbnb's listings features as factors impacting review polarity. The authors found that prices, the host being a super-host, and room type are the main impacting factors on customers' satisfaction. Moreover, in the topic clustering test, positive comments contained "Location" as the most frequent word, while negative comments contained "Check-In". However, Amsterdam as a place has been confirmed that it's affecting Airbnb's customers' satisfaction and could be a motive to generate positive comments. This study was attempting to avoid the biased sample effect, by looking at factors as a percentage of each factor polarity out of the same polarity total in visualization. Regarding this attempt, there are some limitations to this study. The use of one sample such as Amsterdam could affect the result since there are limitations with a non-random sample.

This study highlights some possible orientations for future research. It would be beneficial to compare the results of different countries, to distinguish common factors versus cultural factors. Accordingly, this will help to confirm whether customers' satisfaction is affected by a country's general environment or not. Furthermore, it would be interesting to test text attributes such as room description and host "about" against polarity. Lastly, this study introduced a new method of determining factors affecting customers' satisfaction, by examining factors against polarity using sentiment analysis and regression test, then confirm the results using visualization. This approach can be applied to other studies in hospitality and beyond.

REFERENCES

- [1] Georgiadis L, Iosifidis G, Tassioulas L. On the Efficiency of Sharing Economy Networks. *IEEE Trans Netw Sci Eng* 2019;4697:1–1. <https://doi.org/10.1109/tNSE.2019.2904431>.
- [2] Ju Y, Back KJ, Choi Y, Lee JS. Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *Int J Hosp Manag* 2019; 77:342–52. <https://doi.org/10.1016/j.ijhm.2018.07.014>.
- [3] Sigala M, Rahimi R, Thelwall M. Big data and innovation in tourism, travel, and hospitality: Managerial approaches, techniques, and applications. *Big Data Innov Tour Travel Hosp Manag Approaches, Tech Appl* 2019;1–223. <https://doi.org/10.1007/978-981-13-6339-9>.
- [4] Cui L, Hou Y, Gao M, Yang Y. Exploring the influencing factors of sharing economy sustainability based on a two-mode social network analysis. *2019 16th Int Conf Serv Syst Serv Manag ICSSSM 2019* 2019;1–5. <https://doi.org/10.1109/ICSSSM.2019.8887647>.
- [5] Martin-Fuentes E, Fernandez C, Mateu C, Marine-Roig E. Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *Int J Hosp Manag* 2018;69:75–83. <https://doi.org/10.1016/j.ijhm.2017.10.016>.
- [6] Blal I, Singal M, Templin J. Airbnb's effect on hotel sales growth. *Int J Hosp Manag* 2018;73:85–92. <https://doi.org/10.1016/j.ijhm.2018.02.006>.
- [7] Guttentag DA, Smith SLJ. Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *Int J Hosp Manag* 2017;64:1–10. <https://doi.org/10.1016/j.ijhm.2017.02.003>.
- [8] Xie KL, Kwok L. The effects of Airbnb's price positioning on hotel performance. *Int J Hosp Manag* 2017;67:174–84. <https://doi.org/10.1016/j.ijhm.2017.08.011>.
- [9] Lin PMC, Fan DXF, Zhang HQ, Lau C. Spend less and experience more: Understanding tourists' social contact in the Airbnb context. *Int J Hosp Manag* 2019;83:65–73. <https://doi.org/10.1016/j.ijhm.2019.04.007>.
- [10] Moon H, Miao L, Hanks L, Line ND. Peer-to-peer interactions: Perspectives of Airbnb guests and hosts. *Int J Hosp Manag* 2019;77:405–14. <https://doi.org/10.1016/j.ijhm.2018.08.004>.
- [11] Su N, Mattila AS. 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* 2019. <https://doi.org/10.1016/j.ijhm.2019.102405>.
- [12] Liang S, Schuckert M, Law R, Chen CC. The importance of marketer-generated content to peer-to-peer property rental platforms: Evidence from Airbnb. *Int J Hosp Manag* 2020;84:102329. <https://doi.org/10.1016/j.ijhm.2019.102329>.
- [13] Liu SQ, Mattila AS. Airbnb: Online targeted advertising, sense of power, and consumer decisions. *Int J Hosp Manag* 2017;60:33–41. <https://doi.org/10.1016/j.ijhm.2016.09.012>.
- [14] Falk M, Larpin B, Scaglione M. The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. *Int J Hosp Manag* 2019; 83:132–40. <https://doi.org/10.1016/j.ijhm.2019.04.023>.
- [15] Yang SB, Lee K, Lee H, Koo C. In Airbnb we trust: Understanding consumers' trust-attachment building mechanisms in the sharing economy. *Int J Hosp Manag* 2019;83:198–209. <https://doi.org/10.1016/j.ijhm.2018.10.016>.
- [16] Wang C (Renee), Jeong M. What makes you choose Airbnb again? An examination of users' perceptions toward the website and their stay. *Int J Hosp Manag* 2018; 74:162–70. <https://doi.org/10.1016/j.ijhm.2018.04.006>.
- [17] Cheng M, Jin X. What do Airbnb users care about? An analysis of online review comments. *Int J Hosp Manag* 2019;76:58–70. <https://doi.org/10.1016/j.ijhm.2018.04.004>.
- [18] Luo Y, Tang R (Liang). Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *Int J Hosp Manag* 2019;80:144–54. <https://doi.org/10.1016/j.ijhm.2019.02.008>.
- [19] Zhang M, Luo X, Cheng X, Fu S. An empirical business study on service providers' satisfaction in sharing economy. *Proc - 2016 IEEE 1st Int Conf Data Sci Cyberspace, DSC 2016* 2017:514–9. <https://doi.org/10.1109/DSC.2016.115>.
- [20] Hawlitschek F, Teubner T, Gimpel H. Understanding the sharing economy - Drivers and impediments for participation in peer-to-peer rental. *Proc Annu Hawaii Int Conf Syst Sci* 2016;2016-March:4782–91. <https://doi.org/10.1109/HICSS.2016.593>.
- [21] Alamsyah A, Rochmah WY, Nugroho DDA. Understanding Public Opinion towards New Sharing Economy Business Model Using Content Analysis. *Proc 2019 Int Conf Inf Manag Technol ICIMTech 2019* 2019;1:300–4. <https://doi.org/10.1109/ICIMTech.2019.8843779>.
- [22] Sari R, Lisanti Y, Luhukay D. The Evaluation of Impact Sharing Economy Critical Success Factor Implementation (case study: Online hospitality). *Proc 2019 Int Conf Inf Manag Technol ICIMTech 2019* 2019;1:455–60. <https://doi.org/10.1109/ICIMTech.2019.8843746>.
- [23] Ch C, Gupta D. Factors influencing customer satisfaction with usage of shopping apps in India. *RTEICT 2017 - 2nd IEEE Int Conf Recent Trends Electron Inf Commun Technol Proc* 2017;2018-Janua:1483–6. <https://doi.org/10.1109/RTEICT.2017.8256844>.
- [24] Kumar V, Reinartz W. Applications of CRM in B2B and B2C Scenarios Part I. 2018. https://doi.org/10.1007/978-3-662-55381-7_16.
- [25] Zhao Y, Xu X, Wang M. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *Int J Hosp Manag* 2019;76:111–21. <https://doi.org/10.1016/j.ijhm.2018.03.017>.
- [26] Lutz C, Newlands G. Consumer segmentation within the sharing economy: The case of Airbnb. *J Bus Res* 2018;88:187–96. <https://doi.org/10.1016/j.jbusres.2018.03.019>.
- [27] Edelman BG, Luca M. Digital Discrimination: The Case of Airbnb.com. *SSRN Electron J* 2014. <https://doi.org/10.2139/ssrn.2377353>.
- [28] Guo L, Li J, Wu J, Chang W, Wu J. A Novel Airbnb Matching Scheme in Shared Economy Using Confidence and Prediction Uncertainty Analysis. *IEEE Access* 2018;6:10320–31. <https://doi.org/10.1109/ACCESS.2018.2801810>.
- [29] Nguyen LS, Ruiz-Correa S, Mast MS, Gatica-Perez D. Check Out This Place: Inferring Ambiance from Airbnb Photos. *IEEE Trans Multimed* 2018;20:1499–511. <https://doi.org/10.1109/TMM.2017.2769444>.
- [30] Situmorang KM, Hidayanto AN, Wicaksono AF, Yuliatwati A. Analysis on customer satisfaction dimensions in peer-to-peer accommodation using latent dirichlet allocation: A case study of airbnb. *Int Conf Electr Eng Comput Sci Informatics* 2018;2018-October:542–7. <https://doi.org/10.1109/EECSI.2018.8752912>.
- [31] Park E. Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches. *J Retail Consum Serv* 2019;51:14–8. <https://doi.org/10.1016/j.jretconser.2019.05.019>.
- [32] Cuquet M, Fensel A, Bigagli L. A European research roadmap for optimizing societal impact of big data on environment and energy efficiency. *GIoTS 2017 - Glob Internet Things Summit, Proc* 2017. <https://doi.org/10.1109/GIOTS.2017.8016274>.
- [33] Inside Airbnb n.d. <http://insideairbnb.com/>.
- [34] Li J, Lowe D, Wayment L, Huang Q. Text mining datasets of β -hydroxybutyrate (BHB) supplement products' consumer online reviews. *Data Br* 2020;30. <https://doi.org/10.1016/j.dib.2020.105385>.
- [35] Osborne JW. Simple Linear Models With Polytomous Categorical Dependent Variables: Multinomial and Ordinal Logistic Regression. *Regres Linear Model Best Pract Mod Methods* 2020:133–56. <https://doi.org/10.4135/9781071802724.n6>.
- [36] Chattopadhyay M, Mitra SK. Do airbnb host listing attributes influence room pricing homogeneously? *Int J Hosp Manag* 2019;81:54–64. <https://doi.org/10.1016/j.ijhm.2019.03.008>.
- [37] The Economist Intelligence Unit. Safe Cities Index 2019 2019.