

Article

A Comparative Automated Text Analysis of Airbnb Reviews in Hong Kong and Singapore Using Latent Dirichlet Allocation

Kiattipoom Kiatkawsin ¹, Ian Sutherland ¹ and Jin-Young Kim ^{2,*}

¹ Tourism Industry Data Analytics Lab (TIDAL), Department of Hospitality and Tourism Management, Sejong University, Seoul 05006, Korea; kiatkawsin@gmail.com (K.K.); ianosutherland@gmail.com (I.S.)

² Department of Hotel and Tourism Management, Dong Seoul University, Gyeonggi-do 13117, Korea

* Correspondence: fiu9882@naver.com

Received: 20 July 2020; Accepted: 17 August 2020; Published: 18 August 2020



Abstract: Airbnb has emerged as a platform where unique accommodation options can be found. Due to the uniqueness of each accommodation unit and host combination, each listing offers a one-of-a-kind experience. As consumers increasingly rely on text reviews of other customers, managers are also increasingly gaining insight from customer reviews. Thus, this present study aimed to extract those insights from reviews using latent Dirichlet allocation, an unsupervised type of topic modeling that extracts latent discussion topics from text data. Findings of Hong Kong's 185,695 and Singapore's 93,571 Airbnb reviews, two long-term rival destinations, were compared. Hong Kong produced 12 total topics that can be categorized into four distinct groups whereas Singapore's optimal number of topics was only five. Topics produced from both destinations covered the same range of attributes, but Hong Kong's 12 topics provide a greater degree of precision to formulate managerial recommendations. While many topics are similar to established hotel attributes, topics related to the host and listing management are unique to the Airbnb experience. The findings also revealed keywords used when evaluating the experience that provide more insight beyond typical numeric ratings.

Keywords: automated text analysis; latent Dirichlet allocation (LDA); Airbnb; online reviews; Hong Kong; Singapore

1. Introduction

When Airbnb was launched, it was considered as a safer and more exciting alternative to couch surfing, so most industry experts and global hotel chains did not take much notice [1,2]. The threat was not apparent at first, but 12 years later, Airbnb is arguably the world's largest accommodation provider [3]. Among its competitors in the informal accommodation sector, also known as peer-to-peer accommodation, Airbnb has been the most successful. Together with other companies such as Uber (ride-sharing), Lime (bike rental), JustPark (car parking space) among many others representing the sharing economy, Airbnb takes most of the headlines [4]. What most of these companies have in common is that the idea began as being so radical, that mainstream consumers did not find their attributes to be appealing. Over time, these alternative attributes transformed the market and they became the new normal. Airbnb now competes in most levels of the accommodation market, from low-end to luxury, solo to large groups, and accommodates both leisure and business travelers [2,5]. This life-cycle is typical of disruptive innovations [1].

The industry has reached a point where traditional methods are not able to digest the tremendous amount of data being produced efficiently [5,6]. Likewise, researchers and marketers also need to

look for disruptive tools and techniques to transform the data into knowledge and insights. Customer open-ended textual reviews contain a wealth of information that could potentially be a new source of business intelligence [7]. In the last few years, a handful of research studies have attempted to analyze text data and determine the overall sentiment of tourist reviews [8–12]. Most sentiment analysis has been done using a lexicon- or dictionary-based approach, where words are coded with either positive or negative sentiment, then frequencies of words are counted that appear in a text, or set of texts, to determine the overall sentiment [13,14]. However, this approach only considers words in isolation. Moreover, numeric ratings have already been used to determine the overall sentiment and satisfaction [15]. Hence, a new family of automated text analysis based on word co-occurrences was adopted in order to determine the topics of discussion in text corpora. One of these techniques is latent Dirichlet allocation (LDA). LDA is an unsupervised topic modeling technique that is used to identify and extract topics from a large amount of text [16]. Having a large volume of text reduced into a smaller number of topics with important keywords helps highlight important issues being discussed in the data, and it is particularly useful for gaining insight into customers' online reviews.

Previous studies have used LDA to analyze one million Airbnb reviews and found 43 topics in four groups [17]. While, another study analyzed over 120,000 accommodations in Korea to extract 12 topics [18]. A third study utilized 266,544 hotel reviews on TripAdvisor to find 30 topics of interest [19]. When compared to deductive studies that were designed to test hypotheses based on existing theories, the data-driven topic modeling approach is known to be able to distinguish between topics more precisely, thereby providing more subtle and detailed insights to managers. For example, international hotel guests in Korea have been found to discuss accommodation location in three distinctive topics: accessibility to the accommodation from outside of the city, mobility within the city, and points-of-interest in the surrounding neighborhood where the accommodation is located [18]. However, in most existing theories location would only be measured as one dimension.

Nevertheless, LDA still needs further validation and no previous studies have attempted to compare two sets of reviews in the context of tourism and hospitality [20]. Unlike using the same conceptual framework and validating it using two separate datasets, the study compares two sets of reviews by two topic extraction processes to help discover topics that might be unique to the destination and its attributes. Specifically, each dataset is subjected to two separate processes for estimating the optimal number of topics before the number of topic parameters are set for each of the topic extraction processes. This specific type of comparison allowed the topic dimensions of each destination to be freely extracted independently of one another. Subsequently, the content of each topic is compared and discussed. Hence, the approach allowed for comparisons at two stages, number of topics, and extracted content.

This present study aims to fulfill the lack of comparative studies in topic modeling research by comparing topics extracted from guest reviews of Airbnb in Hong Kong to those in Singapore. Both destinations are comparable in many aspects and have long been considered rival destinations [21–24]. Specifically, the study first gathers Airbnb text reviews from properties in both destinations. After data screening, the optimal number of topics is calculated. Determining the number of topics is necessary to set the parameter before subjecting the data for topic extraction using LDA. Upon the completion of topic extraction, each topic and its associated keywords are analyzed and labeled. More importantly, a second validation process is performed to further validate the topic labels using the top comments for each topic. In the next sections, a thorough review of relevant literature is discussed. Then, the methodology is explained followed by the results of topic extraction and topic labeling. Lastly, discussions of the results along with the implications of the studies can be found in the last section.

2. Literature Review

2.1. Airbnb Experience

Airbnb has been one of the main drivers in the emerging concept of smart tourism [25]. In its simplest form, Airbnb acts as a mediator connecting hosts with rooms for rent to travelers looking to rent rooms [1]. The critical role of Airbnb is then to create and manage the marketplace where hosts can list and guests can discover unique accommodations from around the world [1,5,26]. Airbnb has often been used as a case study of disruptive innovation that has occurred in recent years [1]. The innovative nature of Airbnb that has disrupted the lodging industry was not simply down to its ability to offer price competitive options, but it entirely changed the way tourists view, select, and experience accommodations while traveling [1,2,27–29]. Additionally, the disruption has come not only from the demand side but also from the supply side. People with rooms to rent can now list their units on Airbnb without much investment and prior experience [30,31]. The type of accommodation that would never be available to the general public became readily available [32]. This disruption to the lodging paradigm has triggered a plethora of research to examine the attributes of peer-to-peer accommodations [17,27–29,33,34]. Most notably, research projects comparing the experiences of hotel and Airbnb guests has received much research attention [2,32,35].

In general, the Airbnb experience shares many similarities with hotels. Research has found that when choosing hotels, nine categories of attributes are found to be important. They are services, the physical environment, location, room, price/value, food and beverage, brand image, safety and security, and marketing [36]. Another study identified seven attributes that are equally important in online reviews for both hotels and Airbnb. The seven attributes are cleanliness, safety and security, location, amenities, communication, services, and decoration [35]. In research using automated text analysis techniques, location and amenities have also been found to be important topics of discussion among online reviews of Airbnb properties in Sydney, Australia [26]. The results imply that some basics of hotel accommodations are still expected when choosing and staying at Airbnb properties. Moreover, hotels have been found to still hold some advantages over peer-to-peer accommodations, such as security, hygiene level, and standardized quality [37–39].

Unsurprisingly, some research has focused on identifying the distinguishing characteristics of the Airbnb experience that differ from traditional accommodations. Distinctive attributes of Airbnb when compared with hotels were pets, atmosphere, flexibility, value for money, quality assurance, the need to clean the place before check-out, pressure to write good reviews, sharing room amenities with strangers, and accessibility for disabled people [35]. A study of American and Finnish peer-to-peer guests found two distinctive factors that drive the use of such accommodation options: social appeal, and economic appeal [40]. Social appeal refers to the desire to interact with the host who is most likely to be a local. Thus, guest–host interactions provide the opportunity for guests to ask for recommendations and guidance. On the other hand, the economic appeal is simply the price competitiveness that Airbnb offers when compared to hotels. Authenticity was found to be another significant contributor to the unique Airbnb experience. The exposure to local hosts, staying at a local residential home, and getting a local’s recommendations on local attractions helps form the perception of authenticity, which is found to offer that unique experience appeal [41–43].

Both social appeal and authenticity highlighted the integral role of the host in fostering the Airbnb experience. In contrast to hotels where service interactions are often faceless, Airbnb guests usually develop a much closer relationship with the hosts. The interaction with the host starts from the information search stage because unit descriptions are usually written by the host. Bookings are then confirmed by the host, and all subsequent communications will also be with the same person. Previous research has also been found to empirically support the role of hosts in online text reviews. An analysis of over one million Airbnb reviews in New York City found that hosts were prominently featured as a topic of discussion [17]. Similarly, the host was one of the four themes among Airbnb reviews of properties in Sydney [26]. In contrast, reviews analyzed using topic modeling of hotel accommodations

on three prominent online travel agencies did not find a host to be among the topics of discussion [18]. Likewise, reviews on TripAdvisor of various hospitality sectors also yielded no topic related hosts [19]. Hence, Airbnb hosts are both critical and unique in the peer-to-peer accommodation experience.

2.2. Online Text Reviews

One of the key elements fueling the transition from eTourism to Smart Tourism was the individuals' drive to personalize and co-create the tourism experience [44,45]. Tourists now seek personalized travel experiences that are tailored to fit their exact needs and wants [46]. Thus, more and more tourists rely on standard or common itineraries less than before. Instead, they search for the experience they personally desire and co-create their own travel itineraries [47]. Airbnb's diverse accommodation types and options are critical attributions to the smart tourism paradigm. The wide range of accommodation units, most of which are unique, has the potential to fulfill anyone's imagination. However, the range and uniqueness of each listing also entails a lack of standardization and transferability of previous experiences [2]. In other words, only the review content of the exact same property (to some extent, the same host) is applicable and useful for prospective guests. Hence, reliance on guests writing and sharing their experience in the form of written text reviews has become paramount.

Now, reviews are fundamental to the success of Airbnb, and the overall smart tourism ecosystem [6]. Prospective guests seeking to make the optimal accommodation choice have to gather the relevant information necessary from comments and reviews by previous guests [25]. Hence, the pressure to write reviews after staying is higher compared to staying at hotels [35]. Unlike numeric ratings, where the scores can be easily understood and compared, albeit subjected to the various amount of biases [15], text reviews offer a richer and deeper set of information. Text reviews are usually presented in an unstructured manner creating the inherent issues of (1) making it challenging to locate relevant information, also known as uncertainty, and (2) creating confusion when there is conflicting information, or equivocality [25]. Increasing the number of reviews can help decrease the degree of uncertainty while users can learn to recognize which information to trust. In other words, the participation in writing and sharing travel experiences is a core component in not just enhancing the Airbnb ecosystem, but the overall smart tourism ecosystem [6,25,46].

2.3. Automated Text Analysis

There are predominantly three approaches used for automating text analysis tasks today [20]. The first is to use a dictionary-based approach. Users predefine words with topics, emotions, and other desired properties. Then, words within the documents are detected and defined properties can help produce the output [13,14]. This approach is a popular method for sentiment analysis [8,9], where words such as 'great' and 'clean' might be defined as positive whereas 'late' and 'dirty' might be negative. When there are more positive words than negatives within a document or a review, positive sentiment is assumed for that particular review. There are already developed dictionaries ready to be used, and users have the freedom to add more properties and words to such dictionaries. However, this approach is non-contextual and only counts the words in isolation in what is known as a bag-of-words model. Thus, the dictionary-based approach does not have the ability to capture sarcasm, metaphors, or idioms [20].

The second approach, feature extraction, is similar to the first, the difference lies in the use of machine learning algorithms to define the features [20]. The process requires training of an algorithm and after it has been trained, it should be able to detect features from a text corpora [10,20]. The feature extraction approach can be used in sentiment analysis, gender detection, native/non-native writer detection, and in other applications [11,48]. While models can be trained without human assistance, these models require large datasets with known features to learn from. Additionally, a trained algorithm in a specific domain is unlikely to be useful when analyzing texts from a different domain [20,48]. In hospitality and tourism studies, sentiment analysis has been the most popular application in automated text analysis [8,9,12]. However, review sentiments can be redundant

due to most review platforms (such as Airbnb, TripAdvisor, and online travel agencies) requiring numeric ratings, which can also serve as an accurate representation of the overall sentiment [15,18,49]. In addition, a single written review often consists of different topics and mixtures of sentiments [16]. Thus, assuming the overall sentiment of a review based on the majority proportion of sentiment can be misleading. More importantly, the first two approaches usually analyze documents based on individual words sometimes with added rules for corrections, but words are rarely used in isolation and setting rules may not always solve contextual issues [20,50].

2.4. Latent Dirichlet Allocation (LDA)

The third and most recent approach is based on word co-occurrences [20]. This approach was developed to understand how words are combined to convey meaning. In other words, the assumption that words do not appear in a sentence randomly is the logic behind the word co-occurrence approach [16,51,52]. Capturing the relations between words can distill the total number of dimensions of the dataset into more manageable numbers, and more importantly, it helps to find the most likely topic structure that is discussed in a text corpora, such as a dataset of online guest reviews [20]. The algorithm utilizes a document-term matrix to map the frequency of word co-occurrences [16,53]. This approach is also referred to as topic modeling since the output is the latent topic structures within a dataset. Topic modeling is a branch of natural language processing under the larger artificial intelligence umbrella [54]. Among the topic modeling techniques, LDA is the most widely used today [18]. Its recent popularity among automated text analytic projects could be attributed to how this approach produces results that are most useful when analyzing a large amount of text data [17–19,54].

The extracted topics can help highlight which topics are important among reviewers. Text reviews are unstructured, and reviewers would mostly write what was memorable, either positive or negative. Hence, if a topic is not relevant, reviewers would not write about it often. Consequently, the said topic would not be extracted by the algorithm if it is not discussed frequently. Furthermore, LDA is an inductive approach. The results represent the voice of the customers, rather than confirming an existing theory that may or may not cover all the necessary topics that are relevant in the contemporary context.

Nevertheless, LDA still has its limitations. First, LDA produces results without the awareness of the meaning, albeit unsupervised. Thus, it still requires human interpretation of the topic and for labeling purposes, similar to factor analysis results [16,17]. Second, the algorithm cannot determine any sentiment nor is it able to classify results based on other properties. Therefore, it requires some degree of classification and subsetting of the data before extracting topics [19]. Lastly, Grimmer and Stewart [50] stressed that any quantitative methods to analyze language are wrong because they do not follow the human way of generating and constructing language. Therefore, all quantitative methods are not meant to replace humans but instead are there to augment humans' abilities. Using the examples above, LDA allowed humans to analyze over one million reviews [17,20]. Then, validation of the output is of the utmost importance. These last sets of limitations are not exclusive to LDA but should apply to most, if not all, automated text analysis techniques of today [50].

3. Methods

3.1. Research Settings—Hong Kong and Singapore

As tourism destinations, Hong Kong and Singapore have long been considered rivals and both countries have been compared on many occasions [21–24]. Both international metropolitans offer similar sets of attractions in many areas such as impressive skylines, harbors, culinary varieties, heritage, shopping, and more [21,23,24]. Tourism infrastructure in both cities is also among the worlds' leaders such as well-connected flight routes through their renowned international airports, efficient public transportation systems, and leading of information and technology infrastructures [22,55,56]. The number of international arrivals has also been competitive. Hong Kong reported 55.9 million arrivals between January and December 2019. However, 43.8 million were from Mainland China alone,

leaving only 12.1 million of non-Mainland China arrivals [57]. Singapore recorded 19.1 million arrivals during the same period, which only 3.6 million were from Mainland China [58].

In the Airbnb arena, both have a large number of listings. According to Inside Airbnb, at the time of writing Hong Kong has 9748 listings [59] whereas Singapore has 7975 total listings [60]. Given the similar characteristics of both destinations, a comparison between the two would have some advantages. A comparison would help validate the analysis method (LDA) and increase internal validity. Previous studies have either applied a topic modeling analysis on a single destination [17,18,26,35] or combined data of different destinations but only extract the topic in aggregate [19]. A comparison involves independently calculating the optimal number of topics, then each data set (Hong Kong and Singapore) would be subjected to two separate topic extraction process using LDA. Any differences in the findings should highlight the unique experience at each destination, and vice versa, similarities would imply aspects of competitiveness.

3.2. Data Collection

The raw data was obtained from Inside Airbnb (insideairbnb.com), an independent and non-commercial website providing public data from Airbnb. Inside Airbnb uses a collection of Open Source technologies to mine, process, and visualize Airbnb data. The website aimed to provide the public with free and unbiased insights into Airbnb listings [61]. Although the website provides many facets of data, this present study only focused on the text review written by previous guests. Inside Airbnb also provide data sets of listings in cities across the globe, using the same mining methods. The use of data available on Inside Airbnb has been accepted by previous researchers [17,26,35]. Moreover, available data sets are frequently updated.

The data for Hong Kong was collected as of 29 April 2020. The Singaporean data set was collected on 26 April 2020. At the time of writing, both data sets present the latest available data on Inside Airbnb. The raw data file included the following information listing ID (ID of the accommodation unit), host ID, date of review, reviewer ID, reviewer name, and text review. The raw data for Hong Kong contained a total of 185,695 comments and 93,571 comments from listings in Singapore. Due to the lack of firm structure in text reviews, the raw data were subjected to data screening before analysis.

3.3. Data Screening

Data screening was executed by the authors using R programming language. At this stage, the aim was to remove reviews that have compatibility issues and prepare the texts to be suitable for the LDA algorithm. The screening process follows a similar flow in similar previous studies [62]. First, Google's Compact Language Detector 3 package in R was used to detect and remove non-English reviews. Next, short reviews that contained four words or less are excluded from the analysis. Additionally, automated texts such as cancellation messages were removed. Then, English stop words, or the most common words in the English language, and words that are necessary for grammatical completeness, but not necessary for meaning comprehension were detected and removed. Similarly, punctuations, numbers, emoticons, and programming characters/texts debris from the mining process were removed. Lastly, the *hunspell* package in R was used to stem words and remove proper nouns. Stemming is a process of removing derivatives of a word to retain only the root form. For example, instead of three different words of recommend, recommending, and recommended, stemming will remove -ing and -ed to give three instances of recommend. Proper nouns such as the name of the host, city names, or names of locations are not relevant in understanding the overarching theme of a topic. Instead, they may cause a confounding effect. Hence, they are recommended to be removed before further analysis. As a result of data screening, the number of reviews retained for further analysis was 104,803 for Hong Kong, and 74,829 for Singapore.

4. Findings

4.1. Optimal Number of Topics

Although LDA is an unsupervised method, it still needs the inputting of parameters [51]. The most crucial parameter that LDA needs to have is the number of topics specified by the user. LDA assumes that a set of documents (i.e., reviews) consists of multiple topics, and each topic consists of multiple words. LDA can find the relationships between words, then allocate them inside the corresponding topics. However, LDA cannot automatically suggest the number of topics within a collection, or a dataset [51,52]. Hence, there is a need to find the optimal number of topics, then set it as a parameter before executing the LDA analysis.

Currently, no single method or approach to extract the number of topics has been universally agreed upon by researchers. Since the introduction of LDA in 2013 [16], several methods have been put forth—mainly there have been four often cited methods for determining the optimal number of topics in a dataset, namely Griffiths and Steyvers (2004) [63], Cao et al. (2009) [52], Arun et al. (2010) [53], and Deveaud et al. (2014) [51]. Prior research using LDA to analyze tourism-related online reviews have adopted different approaches. Specifically, a study of online hotel reviews in 16 countries used Griffiths and Steyvers's approach [19]. However, in two studies, one on online Airbnb accommodation reviews of properties in New York City [17] and another about hotel reviews on three online travel agencies [18], the determined number of topics used Deveaud et al.'s method, while also considering its validity with Cao et al.'s method. By normalizing the scores of each method on a scale ranging from 0 to 1, each possible number of topics between 2 and 50 was evaluated via grid search. According to Cao et al., the minimum score is most suitable for the number of topics in a given dataset. On the other hand, Deveaud et al. choose the optimal number of topics that maximizes their score the more suitable number of topics would be. We used Deveaud et al. to determine the optimal number of topics for the Hong Kong data set as determined at 12 (see Figure 1) and Singapore at 5 (see Figure 2).

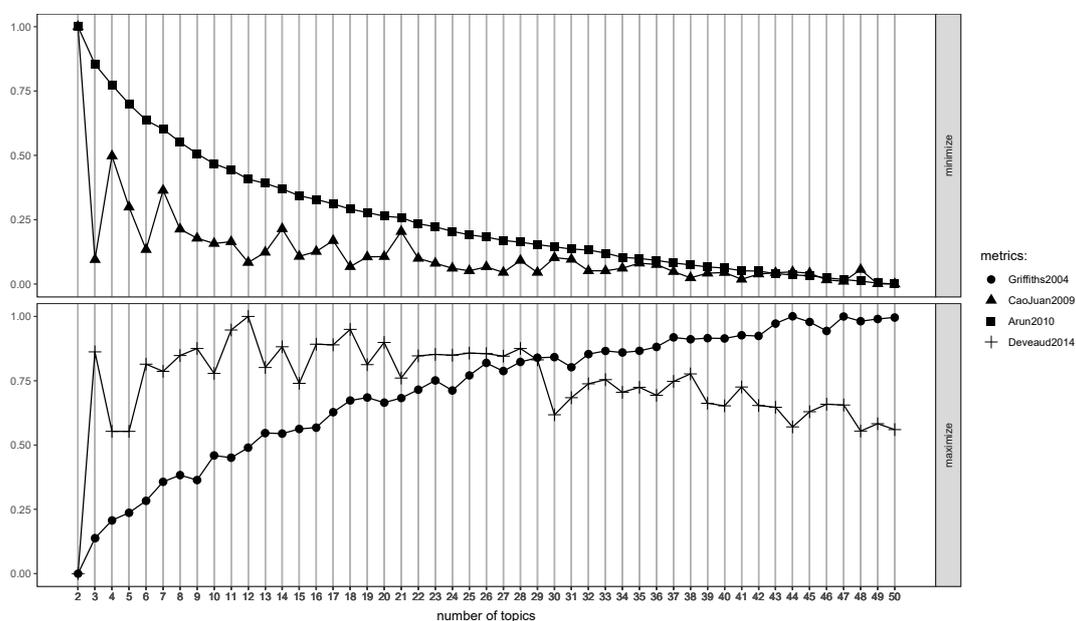


Figure 1. Optimal number of topics comparison chart (Hong Kong).

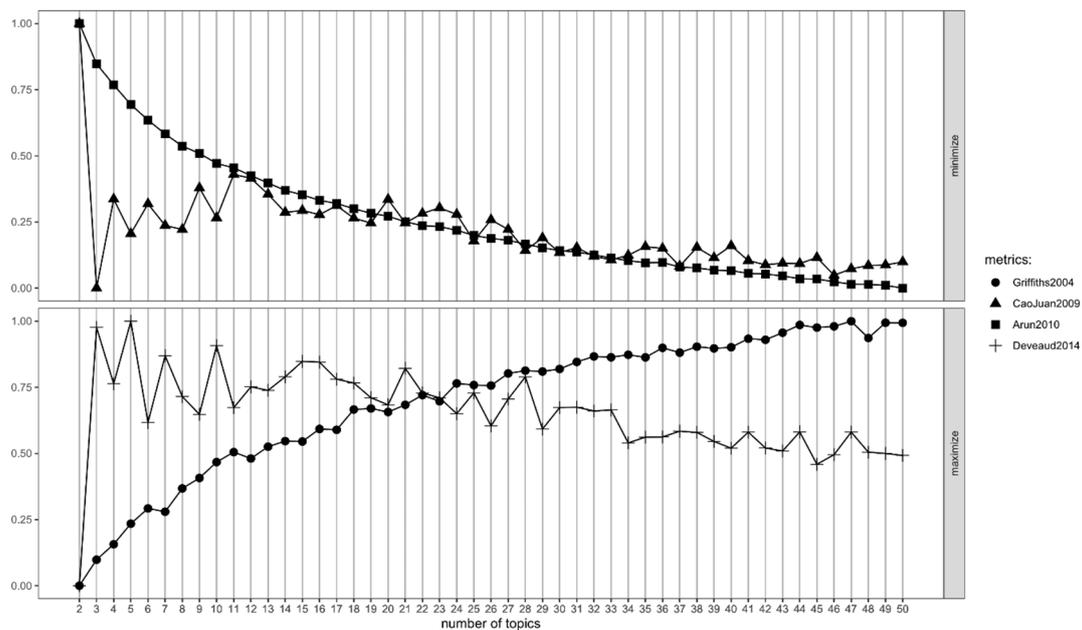


Figure 2. Optimal number of topics comparison chart (Singapore).

4.2. Latent Topics

After the optimal number of topics had been set, the latent Dirichlet allocation algorithm extracted and assigned probabilities to the words for being in each topic. The words with the highest probability (i.e., phi values), are denoted as keywords for each topic in this manuscript. The admixture model implies that the same words can belong in more than one topic. Hence, the words in each topic are organized by the words with the highest probabilities [18]. However, LDA does not explicitly provide the topic labels. Thus, similar to the interpretation of exploratory factor analysis, the researchers are required to interpret the overarching theme of the keywords in each topic and provide a label. In the first step, the top 30 keywords are examined, and labels are attributed to the latent topics. Afterwards, as a second validation process, the document (i.e., online review) with the highest proportion of a given topic is further assessed to validate the given topic label.

4.2.1. Latent Topics of Hong Kong Reviews

A thorough examination of keywords in each of the 12 topics for Hong Kong illustrates clear and distinctive content. The first group of topics is related to the listed accommodation unit. The accommodation unit refers to the core product that the guest purchased through Airbnb. Topic HK1, topic HK2, and topic HK3 are all represented by keywords that describe the unit amenities, unit condition, and accessibility. Unit amenities highlight the guest discussion on what was included at the unit they had been staying in. Thus, the keywords were mainly related to the basic amenities of a unit such as ‘provide’, ‘water’, ‘shower’, ‘towel’, ‘cook’, ‘available’, etc. For the second topic, unit condition, describes the condition of the unit. In this topic, many words were adjectives (‘issue’, ‘noisy’, ‘problem’, ‘bad’, ‘dirty’) that could have been used to describe the condition of nouns (‘building’, ‘door’, ‘toilet’, ‘shower’, ‘sleep’, ‘security’) within the same topic. The third topic is labeled accessibility to the unit, and is reflected in the keywords that are related to how a guest can find the unit. Most of the keywords indicate the explanation and instructions provided for the guests, such as ‘easy’, ‘location’, ‘communicate’, ‘find’, ‘instruction’, ‘clear’, and ‘smooth’, etc.

The second group of topics is related to the location of the unit and the proximity of the unit to its surrounding area. Three topics were identified as part of this topic group and they were labeled accordingly. Topic HK4 is labeled as unit location. Examples of the top keywords contained in this topic are ‘love’, ‘flat’, ‘view’, ‘beautiful’, ‘island’, ‘explore’, ‘neighborhood’, and ‘busy’. These keywords

not only describe the location of the unit concerning the views from the unit but also reflects the nature of Hong Kong. For example, the word ‘island’ would likely explain how the guests could see Hong Kong island from the unit on the Kowloon side. Topic HK5 and topic HK6 are both related to the proximity of the unit to its surrounding area. Specifically, topic HK5 contains words related to attractions, such as ‘restaurant’, ‘shopping’, ‘close’, ‘food’, ‘nearby’, ‘market’, ‘local’, ‘convenience’, ‘plenty’, and ‘temple’. Then, topic HK6 contains words explaining the proximity of the unit in relation to transportation. Thus, the keywords contained in this topic include ‘walk’, ‘station’, ‘minute’, ‘bus’, ‘airport’, ‘close’, ‘distance’, ‘subway’, ‘short’, ‘metro’, and more. The difference between topic HK4 and both topics on proximity are that topic HK4 focuses on the assessment of the unit’s location whereas both proximity topics describe the distance from the unit to local shops and transportation.

The third group of topics reflects the management of the unit. The first topic in this group, topic HK7, is dubbed listing management and it explains guest reviews discussing the arrangement and handling of the listing as reflected in the following example keywords, ‘check’, ‘time’, ‘book’, ‘night’, ‘arrive’, ‘late’, ‘hour’, ‘key’, and ‘luggage’. Topic HK8, host communication, encompasses the following words, ‘host’, ‘quick’, ‘helpful’, ‘responsive’, ‘accommodate’, ‘respond’, ‘question’, ‘fast’, and others. The selected keywords here indicate that the topic revolves around the communication between the host and guests. Both topics are related to the process of booking and other administrative elements of a typical Airbnb booking.

The last group of topics relates to the various dimensions of evaluation. Evaluation refers to the expression of satisfaction in different forms. Specifically, the first topic in this group, topic HK9 is labeled affective evaluation. This topic includes many words that are used to describe emotional responses when evaluating an experience, such as ‘experience’, ‘feel’, ‘friendly’, ‘wonderful’, ‘enjoy’, ‘kind’, ‘warm’, ‘love’, ‘happy’, and ‘safe’. The next topic, topic HK10, is labeled as perceived value. The keywords under this topic have the same underlying theme relating to the price and value of their experience. Thus, the following example keywords were extracted under this topic: ‘small’, ‘space’, ‘price’, ‘expect’, ‘big’, ‘size’, ‘hotel’, ‘location’, ‘pretty’, ‘money’, and ‘worth’. Topic HK11 is named evaluation of host as the keywords relate to the assessment of the host. For example, the keywords in this topic were ‘convenient’, ‘host’, ‘friendly’, ‘house’, ‘comfortable’, ‘helpful’, ‘tidy’, ‘kind’, and ‘staff’. Lastly, topic HK12 is labeled overall evaluation since this topic contains words related to customers’ (dis)satisfaction with the experience. Specifically, some of the top keywords are, ‘recommendation’, ‘high’, ‘clean’, ‘definite’, ‘awesome’, ‘visit’, ‘back’, and ‘heart’. A summary of all 12 topics, the topic labels, and the top 30 keywords are presented in Table 1.

Table 1. List of latent topics and top 30 keywords (Hong Kong).

	Topic	Top 30 Keywords
HK1	Unit amenities	Provide, water, shower, kitchen, towel, work, hot, air, wash, bedroom, bathroom, cook, live, available, condition, kid, machine, include, basic, bed, dryer, heater, adult, amenity, extra, fridge, drink, cold, family, free
HK2	Unit condition	Building, floor, night, door, toilet, issue, shower, sleep, window, bathroom, noise, open, elevator, stair, inside, work, noisy, problem, light, lift, entrance, lock, bad, smell, hard, security, note, wall, main, dirty
HK3	Accessibility to unit	Apartment, easy, location, check, communicate, find, excellent, access, instruction, clear, comfortable, central, tailored, spacious, smooth, amenity, provide, information, fantastic, directions, process, simple, heart, modern, clean, equipped, rain, breeze, standard, efficient

Table 1. Cont.

	Topic	Top 30 Keywords
HK4	Unit location	Love, flat, view, quiet, city, enjoy, beautiful, island, explore, studio, fantastic, spacious, neighborhood, part, decorate, relax, comfortable, modern, town, cool, busy, absolute, equipped, beach, wonderful, balcony, life, beer, central, top
HK5	Proximity to attractions	Restaurant, shopping, close, area, street, food, nearby, market, local, store, bar, night, public, central, convenience, transport, transportation, eat, plenty, mall, lady, convenient, supermarket, downstairs, option, café, eatery, temple, surround, escalator
HK6	Proximity to transportation	Walk, station, minute, bus, airport, close, distance, stop, central, subway, short, location, metro, bay, train, wan, ferry, taxi, convenient, road, causeway, star, tram, exit, building, easily, front, literal, reach, step
HK7	Listing management	Day, check, time, book, night, arrive, late, hour, guest, flight, eat, hidden, arrival, leave, wait, problem, key, left, find, back, morning, call, luggage, turn, long, found, tact, service, meet, end
HK8	Host communication	Host, location, quick, helpful, clean, responsive, accommodate, respond, question, exact, picture, extreme, fast, response, place, describe, answer, prompt, definite, communicate, convenient, photo, excellent, accurate, amenity, request, query, superb, message, description
HK9	Affective evaluation	Time, experience, home, back, feel, trip, friendly, visit, wonderful, family, definite, enjoy, kind, felt, warm, love, happy, tip, live, travel, hospitality, meet, safe, comfortable, hope, future, choose, book, guy, care
HK10	Perceived value	Small, space, people, price, expect, big, travel, size, person, hotel, location, pretty, money, bathroom, couple, standard, hostel, worth, budget, sleep, cheap, especially, fit, large, single, tiny, suit, area, fine, group
HK11	Evaluation of host	Clean, location, convenient, host, friendly, house, comfortable, helpful, place, locate, tidy, kind, cozy, recommendation, staff, neat, facility, pleasant, service, owner, friend, choice, satisfy, safe, family, strategic, environment, ken, hospitable, maintain
HK12	Overall evaluation	Place, recommendation, high, clean, definite, commend, location, cozy, awesome, accessible, visit, back, spacious, comfy, easily, heart, convenient, spot, comfortable, stylish, eat, afford, money, sparkle, thumb, frank, homey, secure, tom, tourist

4.2.2. Latent Topics of Singapore Reviews

An identical process to identify which keywords belong in each topic was performed on the reviews for properties in Singapore. However, the number of topics was set at five. Overall, the results of the Singapore reviews are similar to the results in Hong Kong in terms of the four topic groups but the number of topics in each group was substantially reduced. Topic SG1 is identical to the first two topics in Hong Kong. The topic was named unit and the top keywords were 'bathroom', 'small', 'shower', 'kitchen', 'space', 'water', 'floor', 'sleep', 'air', 'bedroom', 'toilet', and more. Most keywords are related to the accommodation unit. Unlike Hong Kong's results, Singapore's topic SG1 contains words that are both about what the unit has and its conditions. Hence, the topic was broadly labeled unit without further defining the label.

Similar to how topic SG1 contained words related to unit broadly, topic SG2 consists of keywords that are described as the proximity of the unit to its nearby shops and transportation. Therefore, topic SG2 was labeled location to reflect the following top keywords 'walk', 'close', 'station', 'location', 'bus', 'food', 'area', 'shopping', 'restaurant', 'nearby', 'distance', and 'airport'. When briefly compared to the topics derived from the Hong Kong dataset, a noticeable trend emerges in that the topic label for Singapore is following the topic groups of Hong Kong. The label of topic SG3 also reinforces this argument. Topic SG3 was labeled listing management due to the following example keywords: 'check', 'host', 'day', 'time', 'quick', 'book', 'communicate', 'provide', 'late', 'picture', 'respond', and 'arrive'. Hence, the topic represents guests' discussions related to the management of the booking and communication between them and the host.

Topic SG4 and topic SG5 continued with the same patterns, although both topics can be grouped under the same category of evaluation. Topic SG4 was labeled host as the keywords are mostly related to the host, such as 'place', 'clean', 'host', 'recommendation', 'helpful', 'friendly', 'comfortable', 'excellent', 'staff', 'easy', 'cozy', 'awesome'. Some of the keywords were related to the experience and guest (dis-)satisfaction alongside words related to the host. Hence, this topic was grouped under evaluation alongside topic SG5. Topic SG5, labeled evaluation, contains words that express customers' subjective experiences such as 'love', 'family', 'enjoy', 'home', 'comfortable', 'time', 'experience', 'visit', 'back', 'feel', 'wonderful', 'kind', and 'friendly'. Most of the same words can also be seen in the four topics under evaluation from the Hong Kong reviews, although the difference is that they are all under the same topic for Singapore. A summary of the topic labels and top keywords for Singapore can be seen in Table 2.

Table 2. List of latent topics and top 30 keywords (Singapore).

	Topic	Top 30 Keywords
SG1	Unit	Bathroom, small, night, shower, people, work, kitchen, space, wash, water, floor, sleep, guest, air, door, bedroom, toilet, provide, share, towel, building, area, hostel, light, big, pretty, expect, cook, condition, hot
SG2	Location	Walk, close, station, location, bus, minute, food, easy, area, shopping, restaurant, central, stop, city, nearby, distance, access, local, road, public, mall, orchard, airport, street, transport, short, quiet convenient, eat, store
SG3	Listing management	Apartment, check, host, day, time, quick, book, communicate, provide, accommodate, late, picture, respond, arrive, exact, service, response, question, person, back, pleasant, photo, describe, find, hour, early, fast, list, available, week
SG4	Host	Place, clean, location, host, recommendation, helpful, convenient, friendly, high, comfortable, definite, commend, excellent, staff, spacious, easy, money, amenity, cozy, responsive, price, locate, awesome, accessible, accommodation, tidy, worth, comfy, group, neat
SG5	Evaluation	Love, house, family, enjoy, home, pool, comfortable, time, experience, visit, back, feel, travel, view, wonderful, kind, definite, friendly, quiet beautiful, recommendation, trip, fantastic, big, breakfast, felt, spacious, hospitality, warm, live

4.3. Latent Topics Validation

Following the first stage of labeling the latent topic, it was recommended to conduct an additional validation step to ensure the topic labels are accurately reflecting the topic content. Sutherland et al. [18] recommended analyzing the top comments, the comments with the highest proportion of words belonging in each topic, for consistency with the topic label. Due to the nature of unstructured reviews, most reviews contain more than one topic. Therefore, top comments in each topic were

determined when the majority of a review is represented by the respective topic. Hence, analyzing the top comments implies that if the selected comment is mostly about the label given to the topic, the topic label is valid and accurately reflects the content of the topic.

Starting with the Hong Kong reviews. The top comments for topic HK1, unit amenities, contains many sentences that provide information regarding what type of equipment, facilities, and amenities were present in the unit. For example, the top review stated “CK (name of the host) provided a washing machine with dryer” and “Living room has a comfortable sofa bed, LG LED TV with cable and most important Wi-Fi.” Additionally, another top reviewer wrote, “... had everything we needed on a daily basis such as electronic water pot and cook-top, washer with detergent, flatware set, bowls and plates, dishwashing detergent in the kitchen, body wash, shampoo and conditioner, toothpaste and toothbrush, Q-tips, clean towels (all sizes), floor rug, duster, cleaner, and so on.”

The same process was then performed on the top comments belonging in topic HK2, unit condition. The following excerpts were taken to illustrate the exact discussions related to the unit condition, “Lily’s place is very cozy and really clean, I love that the host provided enough towels, toiletries, and other basic needs” along with comments related to the size of the unit such as “don’t stay with more than nine people, it will be crowded due to the room size being tiny.” Similar to how the keywords were different, examining the full written reviews has also indicated that both topics are different with top reviews from topic 1 mostly focusing on listing the available amenities whereas top comments from topic HK2 discuss the condition of the unit.

The same validation process was completed for the remaining topics. During this process, the top comments were consistent with the labels given to each topic. Thus, no modification of the topic labels was needed. Moreover, the topics derived from reviews in Singapore were subjected to the same validation process. Likewise, the top comments mostly contain discussions that were not different than what the keywords indicated. Thus, no adjustments to the topic labels were made. Although this process did not result in any changes to the original topic labels, this validation stage is necessary to ensure the keywords and topic labels are accurate. However, it is important to emphasize that topic labeling should be done using the keywords as reading top comments first without the keywords would make it difficult to detect and distinguish topics from reviews. This is due to how a review often contains more than one topic discussed, and especially when many topics are unique but still fall under the same group. Hence, LDA allows the precise identification of different topics that humans may have a difficult time distinguishing.

5. Discussion

5.1. General Discussions

Overall, the relatively large amount of data used for analysis should help with the generalizability of the results for each destination. As more data is available for analysis, the more accurate the results are [19]. Additionally, online guest reviews have long been considered a credible reflection of the experience and are usually considered to be free of any commercial biases. In general, the difference between Hong Kong and Singapore reviews already started at the optimal number of topics, where Hong Kong recommends 12 topics whereas Singapore suggests only five. Comparing to previous studies using LDA [17–19,35], the optimal number of topics was higher than the established dimensions found in deductive theories. Thus, Hong Kong’s 12 topics offer a more precise identification of attributes that were important among Airbnb guests. Interestingly, the scope of topics is identical between Hong Kong and Singapore. All topics of each destination could be logically assigned into four groups: unit, location, listing management, and evaluation.

The first group of topics was named the unit. Room is often used in the hotel context, but due to the diverse types of accommodation offered by Airbnb, an accommodation unit can include, a shared room, a private room, or an entire apartment. A study by Dolnicar and Otter [36] found one out of nine hotel attributes that were important to guests was related to the room. Where Hong Kong’s data

produced three distinctive topics that highlight what amenities are available, the conditions of the rooms, and the accessibility to the unit. Even if Singapore's data only produced one topic under the unit, the keywords within each topic should still be able to provide useful indications to both hosts and other guests. In addition, when comparing the same topic between Hong Kong and Singapore, some words are specific to the nature of each location. For example, Hong Kong's proximity to transportation included words such as central, causeway, ferry, and tram. All of which are specific to Hong Kong, Central and Causeway Bay are names of tourist hotspots whereas ferry and tram are modes of public transportation unique to Hong Kong. Likewise, Singapore's topic on location included no words such as ferry and tram but instead, orchard and mall, which are the names of a popular area and shopping malls are common attractions in Singapore. The examples here illustrate the power and potential of topic modeling techniques in providing additional insights into guest experience staying in Airbnb in addition to supporting existing theories, highlighting the advantage of LDA compared to deductive theories that may yield results insensitive to the research local settings.

Topics related to listing management and hosts are unique to peer-to-peer accommodation experiences. While the vast variety of accommodation options is an advantage, the challenge is to establish trustable and accurate listing descriptions. Listing accuracy has been discussed in the context of online travel agencies before but the issues do not feature as prominently among hotel guests [64]. Closely linked to listing management is communication with the host. Specifically, the responsiveness of the cost and flexibility. Superficially, hosts and hotel staff may be thought of as playing the same role. However, in the Airbnb context, interactions with hosts started from the information searching stage until the end of the experience. The relationships are forged over a much longer period and are more in-depth. Thus, it is no surprise that topics related to the hosts featured prominently in many peer-to-peer experience research [27,31,35,40,42]. Consequently, social interaction with local people is closely related to the perception of authenticity. Thus, marketers and hosts could use this element as a key differentiator in their marketing materials.

Another point of discussion here is the topic of evaluation. The topics classified under evaluation are topics that can be found in previous studies: affective evaluation, perceived value, evaluation of host, and overall evaluation. Among these, overall evaluation is commonly referred to as satisfaction. Satisfaction is an important construct that measures the holistic evaluation of the experience. Even though the customers perceived different cues separately, the experience is often remembered holistically [65]. However, satisfaction is often associated with positive sentiment, but topic modeling only focuses on identifying topics of discussions. Thus, the topic is labeled overall evaluation and should be interpreted as having a neutral sentiment. On the other hand, affect has been suggested as a powerful contributor to the overall satisfaction in various hospitality and tourism contexts [66–68]. Thus, the topic affective evaluation extracted from Airbnb reviews showed a level of consistency with prior research findings. The topic of perceived value is also a commonly found construct in prior studies [69,70]. Lastly, the topic of evaluation of hosts has been consistent with previous findings among peer-to-peer accommodation experience, while not sharing this similarity with hotels [17,38,40]. Arguably, the discussions about Airbnb hosts and hotel staff could be considered identical, however, the difference is due to the context and how the communication is established.

The comparison of Airbnb reviews between Hong Kong and Singapore revealed both similarities and differences. The optimal number of topics were different between the two datasets. The difference could be attributed to the variety and content depths of the reviews. Interestingly, the scope of the content is nearly identical despite the different number of topics being extracted. Between the two sets of results, arguably the higher topic dimension offers more specific and deeper insights into Airbnb experiences. The more holistic outputs from the Singaporean reviews offer a more compact and general view of the experience. Yet, the most important findings are how both sets of results are identical in terms of scope. Hence, the lower number of topics does not result in a loss of information. Furthermore, when compared to existing findings of hotel experience, peer-to-peer guests perceived relatively similar cues of attributes and quality, albeit using terminologies that are specific to the context of Airbnb.

5.2. Implications

Peer-to-peer accommodation hosts can learn from the research findings in that, although reviews are written during the post-trip stage, the content and sentiments reflect not only the trip and post-trip, but also the pre-trip as well (e.g., communication with the hosts, etc.). Therefore, any hosts of peer-to-peer accommodation should not only focus on delivering quality service only between checking-in to checking-out but from pre-trip to post-trip. In other words, listing accuracy, comprehensiveness, and completeness are not just important attributes leading to overall satisfaction but can also attract more prospective guests.

The managers of traditional hotels and interested individuals can also benefit from the research findings in a few ways. First, this research used an unsupervised method to analyze a large amount of text data and extract results in a way that can offer both new insights and a monitoring system of guests' experiences. When data from other regions and parts of the world are analyzed, uniqueness and trends may emerge leading to new insights. However, it is important to note that topic modeling is still a relatively new tool being utilized and further advancements in the technology may help offer even more depth and utility. Therefore, results from topic modeling are most useful when combined with other sources of information (such as the numeric rating scores) as it has proved to be a useful tool in augmenting the human ability to dissect a large amount of text data in a relatively efficient manner.

5.3. Limitations and Recommendations for Future Research

Topic modeling is a relatively new method in hospitality and tourism management research. While the outputs offer expanded views into accommodation experience research, the dimensions would greatly benefit from theoretical support. Hence, topic modeling should not be viewed as a replacement method but instead a complementary method. Thus, future research is encouraged to synthesize both methods. A limitation of topic modeling is the inability to determine sentiment. Still, combining topic modeling with other methods and data types should offer even greater insights for marketers and managers alike. Another limitation of LDA is the lack of a proven method to test relationships between topics. Unlike established psychometric theories where the hierarchy of cognition is strongly proven empirically, LDA does not have a proven method to establish hierarchical relationships between the topics. Ward clustering has been used to establish relationships, but the technique has yet to gain widespread acceptance. Future studies are strongly encouraged to explore ways to test relationships among the extracted topics. In general, there are other possibilities to combine topic modeling with other types of data and analysis techniques to further advance its usefulness and broaden its application. Additionally, the results of this study should be viewed with caution as Airbnb does not distinguish between reviews of accommodation or experience if both are offered by the same host. Lastly, most reviews are written in text. However, the option to upload images as part of a customer review has been added to many websites. Future studies may wish to expand the scope to include images as a source of data for analysis.

Author Contributions: Conceptualization, K.K., I.S. and J.-Y.K.; data collection, K.K.; analysis, I.S. and K.K.; writing, K.K.; editing, I.S.; supervision, J.-Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guttentag, D. Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Curr. Issues Tour.* **2015**, *18*, 1192–1217. [[CrossRef](#)]
2. Mody, M.A.; Suess, C.; Lehto, X. The accommodation experiencescape: A comparative assessment of hotels and Airbnb. *Int. J. Contemp. Hosp. Manag.* **2017**, *29*, 2377–2404. [[CrossRef](#)]

3. McRae, H. Facebook, Airbnb, Uber, and the Unstoppable Rise of the Content Non-Generators. Available online: <https://www.independent.co.uk/news/business/comment/hamish-mcrae/facebook-airbnb-uber-and-the-unstoppable-rise-of-the-content-non-generators-10227207.html> (accessed on 16 July 2020).
4. Robinson, D. Sharing Economy Companies List: Featuring Airbnb, Uber, Omni and Hubble. Available online: <https://www.ns-businesshub.com/technology/sharing-economy-companies-list/> (accessed on 16 July 2020).
5. Guttentag, D. Progress on Airbnb: A literature review. *J. Hosp. Tour. Technol.* **2019**, *10*, 814–844. [CrossRef]
6. Kiatkawsin, K.; Sutherland, I.; Lee, S.K. Determinants of Smart Tourist Environmentally Responsible Behavior Using an Extended Norm-Activation Model. *Sustainability* **2020**, *12*, 4934. [CrossRef]
7. He, L.; Han, D.; Zhou, X.; Qu, Z. The Voice of Drug Consumers: Online Textual Review Analysis Using Structural Topic Model. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3648. [CrossRef]
8. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment analysis in tourism: Capitalizing on big data. *J. Travel Res.* **2019**, *58*, 175–191. [CrossRef]
9. Kirilenko, A.P.; Stepchenkova, S.O.; Kim, H.; Li, X. Automated sentiment analysis in tourism: Comparison of approaches. *J. Travel Res.* **2018**, *57*, 1012–1025. [CrossRef]
10. Bakshi, R.K.; Kaur, N.; Kaur, R.; Kaur, G. Opinion Mining and Sentiment Analysis. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 452–455.
11. Hemmatian, F.; Sohrabi, M.K. A survey on classification techniques for opinion mining and sentiment analysis. *Artif. Intell. Rev.* **2019**, *52*, 1495–1545.
12. Zhang, L.; Ghosh, R.; Dekhil, M.; Hsu, M.; Liu, B. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *HP Lab. Tech. Rep. HPL-2011* **2011**, 89.
13. Tausczik, Y.R.; Pennebaker, J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **2010**, *29*, 24–54. [CrossRef]
14. Chung, C.K.; Pennebaker, J.W. Linguistic Inquiry and Word Count (LIWC): Pronounced “Luke,” ... and Other Useful Facts. In *Applied Natural Language Processing: Identification, Investigation and Resolution*; IGI Global: Hershey, PA, USA, 2012; pp. 206–229.
15. Smironva, E.; Kiatkawsin, K.; Lee, S.K.; Kim, J.; Lee, C.-H. Self-selection and non-response biases in customers’ hotel ratings—A comparison of online and offline ratings. *Curr. Issues Tour.* **2020**, *23*, 1191–1204. [CrossRef]
16. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
17. Sutherland, I.; Kiatkawsin, K. Determinants of Guest Experience in Airbnb: A Topic Modeling Approach Using LDA. *Sustainability* **2020**, *12*, 3402. [CrossRef]
18. Sutherland, I.; Sim, Y.; Lee, S.K.; Byun, J.; Kiatkawsin, K. Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation. *Sustainability* **2020**, *12*, 1821. [CrossRef]
19. Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [CrossRef]
20. Iliev, R.; Dehghani, M.; Sagi, E. Automated text analysis in psychology: Methods, applications, and future developments. *Lang. Cogn.* **2015**, *7*, 265–290. [CrossRef]
21. Henderson, J.C. Heritage attractions and tourism development in Asia: A comparative study of Hong Kong and Singapore. *Int. J. Tour. Res.* **2002**, *4*, 337–344. [CrossRef]
22. Horng, J.-S.; Tsai, C.-T. (Simon) Exploring Marketing Strategies for Culinary Tourism in Hong Kong and Singapore. *Asia Pac. J. Tour. Res.* **2012**, *17*, 277–300. [CrossRef]
23. Wong, K.K.F.; Kwan, C. An analysis of the competitive strategies of hotels and travel agents in Hong Kong and Singapore. *Int. J. Contemp. Hosp. Manag.* **2001**, *13*, 293–303. [CrossRef]
24. Yeung, S.; Wong, J.; Ko, E. Preferred shopping destination: Hong Kong versus Singapore. *Int. J. Tour. Res.* **2004**, *6*, 85–96. [CrossRef]
25. Bae, S.J.; Lee, H.; Suh, E.-K.; Suh, K.-S. Shared experience in pretrip and experience sharing in posttrip: A survey of Airbnb users. *Inf. Manag.* **2017**, *54*, 714–727. [CrossRef]
26. Cheng, M.; Jin, X. What do Airbnb users care about? An analysis of online review comments. *Int. J. Hosp. Manag.* **2019**, *76*, 58–70. [CrossRef]
27. Poon, K.Y.; Huang, W.-J. Past experience, traveler personality and tripographics on intention to use Airbnb. *Int. J. Contemp. Hosp. Manag.* **2017**, *29*, 2425–2443. [CrossRef]
28. Lutz, C.; Newlands, G. Consumer segmentation within the sharing economy: The case of Airbnb. *J. Bus. Res.* **2018**, *88*, 187–196. [CrossRef]

29. Gutiérrez, J.; García-Palomares, J.C.; Romanillos, G.; Salas-Olmedo, M.H. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tour. Manag.* **2017**, *62*, 278–291. [[CrossRef](#)]
30. Gunter, U. What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area. *Tour. Manag.* **2018**, *66*, 26–37. [[CrossRef](#)]
31. Liang, S.; Schuckert, M.; Law, R.; Chen, C.-C. Be a “Superhost”: The importance of badge systems for peer-to-peer rental accommodations. *Tour. Manag.* **2017**, *60*, 454–465. [[CrossRef](#)]
32. Guttentag, D.A.; Smith, S.L. Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *Int. J. Hosp. Manag.* **2017**, *64*, 1–10. [[CrossRef](#)]
33. Yannopoulou, N.; Moufahim, M.; Bian, X. User-Generated Brands and Social Media: Couchsurfing and AirBnb. *Contemp. Manag. Res.* **2013**, *9*. [[CrossRef](#)]
34. Bridges, J.; Vásquez, C. If nearly all Airbnb reviews are positive, does that make them meaningless? *Curr. Issues Tour.* **2018**, *21*, 2065–2083. [[CrossRef](#)]
35. Zhang, G.; Cui, R.; Cheng, M.; Zhang, Q.; Li, Z. A comparison of key attributes between peer-to-peer accommodations and hotels using online reviews. *Curr. Issues Tour.* **2020**, *23*, 530–537. [[CrossRef](#)]
36. Dolnicar, S.; Otter, T. Which hotel attributes matter? A review of previous and a framework for future research. *Fac. Commer.* **2003**, *1*, 176–188.
37. Mody, M.; Hanks, L.; Dogru, T. Parallel pathways to brand loyalty: Mapping the consequences of authentic consumption experiences for hotels and Airbnb. *Tour. Manag.* **2019**, *74*, 65–80. [[CrossRef](#)]
38. de Castro Cardoso, D.M. What Reasons Lead Guests to Choose Airbnb and Booking.com to Book Peer to Peer Rentals? A Comparison between the Two Major Booking Platforms. Master’s Thesis, University of Porto, Porto, Portugal, 2018.
39. Lu, L.; Mody, M.; Andajigarmaroudi, S. Exploring guest response towards service failure in home-sharing: Service presence and consumption motivation. *Int. J. Hosp. Manag.* **2020**, *87*, 102498. [[CrossRef](#)]
40. Tussyadiah, I.P.; Pesonen, J. Drivers and barriers of peer-to-peer accommodation stay—an exploratory study with American and Finnish travellers. *Curr. Issues Tour.* **2018**, *21*, 703–720. [[CrossRef](#)]
41. Lalicic, L.; Weismayer, C. The Role of Authenticity in Airbnb Experiences. In *Information and Communication Technologies in Tourism 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 781–794.
42. Liang, L.J.; Choi, H.C.; Joppe, M. Understanding repurchase intention of Airbnb consumers: Perceived authenticity, electronic word-of-mouth, and price sensitivity. *J. Travel Tour. Mark.* **2018**, *35*, 73–89. [[CrossRef](#)]
43. Mody, M.; Hanks, L. Consumption authenticity in the accommodations industry: The keys to brand love and brand loyalty for hotels and Airbnb. *J. Travel Res.* **2020**, *59*, 173–189. [[CrossRef](#)]
44. Vargas-Sánchez, A. Exploring the concept of smart tourist destination. *Enl. Tour. Pathmaking J.* **2016**, *6*.
45. Buhalis, D.; Amaranggana, A. Smart tourism destinations enhancing tourism experience through personalisation of services. In *Information and Communication Technologies in Tourism 2015*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 377–389.
46. Femenia-Serra, F.; Neuhofer, B.; Ivars-Baidal, J.A. Towards a conceptualisation of smart tourists and their role within the smart destination scenario. *Serv. Ind. J.* **2019**, *39*, 109–133. [[CrossRef](#)]
47. Gajdošík, T. Smart tourists as a profiling market segment: Implications for DMOs. *Tour. Econ.* **2019**, *26*, 1042–1062. [[CrossRef](#)]
48. Dave, K.; Lawrence, S.; Pennock, D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 20 May 2003; pp. 519–528.
49. Jia, S. Behind the ratings: Text mining of restaurant customers’ online reviews. *Int. J. Mark. Res.* **2018**, *60*, 561–572. [[CrossRef](#)]
50. Grimmer, J.; Stewart, B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* **2013**, *21*, 267–297. [[CrossRef](#)]
51. Deveaud, R.; SanJuan, E.; Bellot, P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Doc. Numer.* **2014**, *17*, 61–84. [[CrossRef](#)]
52. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [[CrossRef](#)]

53. Arun, R.; Suresh, V.; Madhavan, C.V.; Murthy, M.N. On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 391–402.
54. Jacobi, C.; Van Atteveldt, W.; Welbers, K. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digit. J.* **2016**, *4*, 89–106. [CrossRef]
55. Chiu, H.Y.; Chan, C.-S.; Marafa, L.M. Local perception and preferences in nature tourism in Hong Kong. *Tour. Manag. Perspect.* **2016**, *20*, 87–97. [CrossRef]
56. Lim, C.; Zhu, L.; Koo, T.T. Urban redevelopment and tourism growth: Relationship between tourism infrastructure and international visitor flows. *Int. J. Tour. Res.* **2019**, *21*, 187–196. [CrossRef]
57. Hong Kong Tourism Board Visitor Arrivals to Hong Kong in December 2019. Available online: https://partnet.hkta.com/en/research_statistics/latest_statistics/index.html?gvHelp=1 (accessed on 16 July 2020).
58. Singapore Tourism Board Tourism Sector Performance for Quarter 4 2019. Available online: <https://www.stb.gov.sg/content/stb/en/statistics-and-market-insights/tourism-statistics/quarterly-tourism-performance-report.html> (accessed on 16 July 2020).
59. Inside Airbnb Inside Airbnb: Hong Kong. Adding Data to the Debate. Available online: <http://insideairbnb.com/hong-kong> (accessed on 16 July 2020).
60. Inside Airbnb Inside Airbnb: Singapore. Adding Data to the Debate. Available online: <http://insideairbnb.com/singapore> (accessed on 16 July 2020).
61. Cox, M.; Slee, T. How Airbnb's Data hid the Facts in New York City. Available online: <http://insideairbnb.com/reports/how-airbnbs-data-hid-the-facts-in-new-york-city.pdf> (accessed on 16 July 2020).
62. Tirunillai, S.; Tellis, G.J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *J. Mark. Res.* **2014**, *51*, 463–479. [CrossRef]
63. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef]
64. Park, Y.A.; Gretzel, U.; Sirakaya-Turk, E. Measuring Web Site Quality for Online Travel Agencies. *J. Travel Tour. Mark.* **2007**, *23*, 15–30. [CrossRef]
65. Han, H.; Kiatkawsin, K.; Kim, W. Traveler loyalty and its antecedents in the hotel industry: Impact of continuance commitment. *Int. J. Contemp. Hosp. Manag.* **2019**, *31*, 474–495. [CrossRef]
66. Kiatkawsin, K.; Han, H. An alternative interpretation of attitude and extension of the value–attitude–behavior hierarchy: The destination attributes of Chiang Mai, Thailand. *Asia Pac. J. Tour. Res.* **2017**, *22*, 481–500. [CrossRef]
67. Barsky, J.; Nash, L. Evoking emotion: Affective keys to hotel loyalty. *Cornell Hotel Restaur. Adm. Q.* **2002**, *43*, 39–46. [CrossRef]
68. del Bosque, I.R.; Martín, H.S. Tourist satisfaction a cognitive-affective model. *Ann. Tour. Res.* **2008**, *35*, 551–573. [CrossRef]
69. Kim, S.; Ham, S.; Moon, H.; Chua, B.-L.; Han, H. Experience, brand prestige, perceived value (functional, hedonic, social, and financial), and loyalty among GROCERANT customers. *Int. J. Hosp. Manag.* **2019**, *77*, 169–177. [CrossRef]
70. Chi, X.; Lee, S.K.; Ahn, Y.; Kiatkawsin, K. Tourist-Perceived Quality and Loyalty Intentions towards Rural Tourism in China. *Sustainability* **2020**, *12*, 3614. [CrossRef]

