Visitor’s Experience in Angkor Wat, Cambodia: Evidence from Sentiment and Topic Analysis

Abstract

Angkor Wat in Cambodia is one of the most significant archaeological sites in the world inscribed as a UNESCO World Heritage Site. It is also consistently rated as the top Global Landmark by travelers in TripAdvisor. This study employed big data analytics to examine visitor experience via sentiment and topic analysis, given the popularity and volume of visitors annually. The process involved an examination of 32,394 online reviews about Angkor Wat posted between January 2015 - October 2019 on Tripadvisor’s website. The analysis was conducted for reviews with one, two, and three-star ratings, and those with four and five star ratings separately. Results showed highly skewed positive ratings, while similar positive sentiments dominated the reviews. Topic analysis disclosed that sunset and sunrise experience, attraction structure, guided tours, and temple experience were topics attributed mostly to positive sentiments. Conversely, crowding, persistent selling, clothing style, and expenses were illustrated as negative topics. Such insights are useful for managers at Angkor Wat to develop various marketing and management interventions to effectively manage the site and to optimize visitor experiences.

Keywords: Visitors, Sentiment, Topic, Cultural Heritage, Tourism, Data Analytics

Introduction

The world is extensively interconnected through 4.4 billion internet users, 3.5 billion active social media users, and 3.3 billion mobile social media users (Kemp, 2019). As a result of this high digitalization and rapid development of mobile technologies, there has been a significant surge in the use of online media channels. This has enabled ease in accessibility and dissemination of information among consumers, as well as largely influenced the travel and tourism industry (Garín-Muñoz & Perez-Amaral, 2011). More specifically, travelers have gained enormous leverage as they can influence other travelers' decisions with their personal opinions and comments about their experiences with product and services. Such comments commonly known as user-generated contents (UGC) are generally utilized by travelers who seek information and advice during the research phase of their travel planning process and trip decisions (Cox et al., 2009; Leung et al., 2013; Mendes-Filho et al., 2018).

Various online review websites, along with popular social media platforms such as Facebook and Twitter offer billions of personal opinions, which are crucial sources of information for potential travelers. Among them, TripAdvisor is the largest and most well-known online review platform for the travel and tourism industry, where people share opinions and comments about hotels, attractions, historical sites, landmarks, and cruises (Simeon et al., 2017). According to the New Oxford Economics Study, TripAdvisor has influenced 10.3% of global travel spending and 433 million trips with its page views, reviews and scores in 2017 alone (Worldwide Tourism Economics, 2018). TripAdvisor does not only provide travelers with reviews but also compares prices to identify the lowest price (Tripadvisor.com, 2020a). Collectively, TripAdvisor is a significant influencer in the tourism industry globally based on insights generated from travelers.

An extensive body of tourism research has utilized TripAdvisor predominantly for hotel reviews, and offered insights to practitioners to improve service quality (Barreda & Bilgihan, 2013); examine satisfied and dissatisfied customers (Berezina et al., 2016); and compared data quality between different online platforms relying on multiple data sources (Xiang et al., 2017). Furthermore, past research has also used online reviews to investigate the influence of readability and reviewer characteristics on perceived value (Fang et al., 2016), and to identify perspectives toward a specific tourism product (Pearce & Wu, 2015). However, given the utility of this approach with respect to user generated content reviews, there is a paucity of research that have examined visitors' online reviews of heritage sites. The opportunity to examine this specific sector is timely given the global demand for cultural tourism, and recognition as a tool to attain sustainable development goals by United Nations' World Tourism Organization (UNWTO, 2020).

This study aims to assess the experience dimensions of heritage site visitors through sentiments and topics in TripAdvisor reviews of Angkor Wat, Cambodia. Big data analytical methods were employed as this enables to reveal issues within large volumes of data, and also provide practical input for industries (Guo, Barnes, & Jia, 2017). The reason to choose Angkor Wat as a case study are two-fold; first, it is one of the world's most visited heritage sites; second, it was named as the top Global Landmark, which won TripAdvisor Travelers' Choice Award for two consecutive years (2017 and 2018). This award winners are determined based on both the quantity and quality of members' reviews and ratings for landmarks worldwide, congregated over 12 months (TripAdvisor, 2020b). The study makes significant practical contributions by revealing sentiments derived from online reviews and drawing conclusions based on positive and negative aspects of heritage site visitor experience. Overall, the research implications offer heritage site managers and marketing professionals greater insight into the dimensions of heritage site experience to improve experience quality and increase visitor satisfaction. As one of the exceptional studies analyzing big data on heritage sites, the results can be used to better understand the dynamics of tourist behaviour in cultural trips.

Literature Review

User-Generated Content in the Travel & Tourism Industry

The term user-generated content (UGC) gained popularity among both tourism researchers and practitioners since early 2000s (Xiang et al., 2015). The enhancement of contents and file-sharing applications by the internet enabled the creation and distribution apparatus for UGC (Daugherty, Eastin & Bright, 2008). Thus, the trend shifted from company-induced to consumer-generated content with the evolvement of Web 2.0, which enabled passive internet users to be potential content creators (Van Dijck, 2009; Yoo & Gretzel, 2011). UGC has been defined by three main characteristics; (a) needs to be publicly available - Internet, (b) should involve a certain amount of creative effort, and (c) needs to be created outside of professional practices and routines (OECD, 2007, p.18). UGC encompasses a wide range of content types, that includes texts, photos, videos, audios, and documents (Kaplan & Haenlein, 2010).

UGC has been extensively assessed in the tourism and hospitality literature as most research has examined the influence of type (consumer-written versus system-aggregated) and valence (positive or negative) (S. V. Jin & Phua, 2016). Additionally, information accuracy, relevance, reliability, and expertise of the UGC creator has been also identified as other factors to influence travelers' adoption of content (Ayeh, Au, & Law, 2013; Chung & Koo, 2015; Filieri & McLeay, 2013). Past studies have also examined user motivations to create UGC (Daugherty et al., 2008), self-documentation, sharing, hedonic enjoyment (Wu & Pearce, 2016), assist others, and to gain fame and recognition (Munar & Jacobsen, 2014).

UGC has a significant impact on intention to travel, visitors' attitudes towards a destination (Jalilvand & Samiei, 2012), and traveler perceptions of destination brands (Lim, Chung & Weaver, 2012). Similarly, the influence of UGC has been found on consumers' hotel booking behavior (Del Chiappa, Alarcón-Del-Amo, & Lorenzo-Romero, 2016), and the direct online sales of hotel rooms (Ye et al., 2011). UGC provides a much-needed ability to predict consumer preferences that can be instrumental to offer the right marketing mix to customers to increase satisfaction (Christodoulies, Jevons, & Bonhomme, 2012). Consequently, UGC should be treated as a valuable asset for businesses, destination planners, and national tourism organizations in strategic planning processes (Marine-Roig & Clavé, 2015).

Sentiment and Topic Analysis to Understand Visitor Experience

Social media platforms have become an extensive repository of visitors' unforced sentiments and experiences in the form of their expressions, which aid tourism marketers to understand the real sentiments and experience of tourists. As a result, both scholars and businesses, who aim to mine this abundant information and create meaningful solutions have utilized big data generated through three main sources (i.e., users, devices, and operations). These sources generate three types of data: (1) UGC via texts, photos and videos primarily extracted from social media platforms; (2) Spatial-temporal data gathered from sensor devices; (3) Transaction data acquired from online booking, website use, and web search (Li et al., 2018). The volume of data is so large that the analytical capabilities of traditional qualitative, quantitative, and mixed methods are not enough to collect, analyze, and interpret data (Guo et al., 2017; Wang, Feng, & Dai, 2018). Thus, data mining, web crawling, machine learning, natural language processing, and automated analysis methods have been utilized within the scope of big data analytics (Taecharungroj & Mathayomchan, 2019; Xiang et al., 2017).

Among them, two primary computational tools in tourism and hospitality research are automated sentiment analysis and topic analysis. Sentiment analysis also known as opinion mining is built on the principle that the information provided in a text is either subjective (personal opinions, beliefs, feelings) or objective (facts and evidences) (Alaei, Becken, & Stantic, 2019; Kennedy, 2012). In the scope of sentiment analysis, machine learning algorithms are mostly applied on the individuals' subjective reviews to analyze the sentiment polarity and the valence of textual data and determining the polarity of the given review as "positive" or "negative" categories (Pan, Zhong, & Yang, 2012; Schuckert, Liu, and Law 2015). However, Alaei et. al. (2019) suggested that a third category (neutral) should be added to this binary classification to represent objective reviews which generally do not contain positive and negative words. Sentiment analysis which enables researchers and practitioners to reveal positive, negative and neutral sentiments in consumer reviews is widely utilized in tourism field to understand satisfaction and dissatisfaction of travelers. Moreover, the valence of online reviews (positive or negative) is one of the crucial elements that affect potential consumers and their purchase decisions (Ye et al., 2011).

Topic analysis, which is a form of dimension reduction uses a probabilistic model to reveal the co-occurrence patterns of terms that conform to semantic topics in documents (Crain et al., 2012). Topic modelling involves advanced software and mathematical techniques developed in the fields of natural language processing and data mining (Guo et al., 2017). Latent Dirichlet Allocation (LDA) is the most common method in topic modeling. This approach is used to group several words or terms found in textual data and portray as a document to contain latent topics as it has a multinomial distribution over terms (Wang et al., 2018). LDA assumes that there are latent factors (topics) in the text and classifies the words into specific topics based on similarity between parts in textual data (Anandkumar et al., 2012). Topic analysis assist researchers to analyze large-scale online reviews to identify the underlying dimensions of consumer satisfaction based on similar parts in the textual data.

Big data analytics is utilized by the tourism industry to improve efficiency and deliver high-quality services by evaluation of data related to consumers' lodging stays purchase transactions, and preferences (Song & Liu, 2017). In the academic literature, an increasing number of scholars have applied automated sentiment and topic analysis methods to tourism and hospitality related big data. Majority of studies have analyzed data associated to the lodging industry (e.g., Garcia-Pablos, Cuadros, & Linaza, 2016; Lee, Lee, & Koh, 2019), while emergent topics have been related to destinations (Gkritzali, Gritzalis, & Stavrou, 2018), attractions (Kirilenko et al., 2019), and restaurants (Gan et al., 2017). As evident, research about cultural heritage and big data is very limited (Clarizia et al., 2018; Valdivia et al., 2020), and this study of Angkor Wat serves to offer insights about heritage visitor experience.

Research Design and Methodology

Study Area

The site context is Angkor, located in Cambodia, one of the most critical archaeological establishments in Southeast Asia. With over 400 km^2, it has a tropical forest, expanses of cultivated land, numerous rural communities, and some of the most massive and elaborated architectural structures in history (Winter, 2007). Further, Angkor Archaeological Park includes the remnants of the different capitals of the Khmer Empire from the 9th - 15th century (Albert &

Ringbeck, 2015). Some of the prominent are the famous Temple of Angkor Wat and at Angkor Thom, Bayon Temple with its countless sculptural decorations. Based on these historic resources, Cambodia has built its national identity on cultural heritage, which is its primary source of attraction (Chheang, 2008; Winter, 2007).

Angkor was added to the UNESCO World Heritage Site list in 1992, to preserve and further develop its cultural landscape with monumental and archaeological remains (Hauser-Schäublin, 2011). Within this, the pivotal attraction is the Angkor Wat complex. The meaning and significance of Angkor Wat as an architectural wonder has been embraced at the international level (Carter et al., 2015). The expansive site showcases its archaeological richness (Brigitta, 2011), and has a reputation of being the world's largest religious monument erected before the twentieth century A.D. (Fletcher et al., 2015). Angkor Wat demonstrates cultural, religious, and symbolic ideas, as well as holds extraordinary architectural, archaeological, and artistic meaning for visitors. The popularity is evidenced by nearly 2.6 million tourists annually and generates more than US$100 million in revenues (Cheng Sokhorng, 2019). Angkor Wat has been consistently ranked as the top landmark by TripAdvisor reviewers.

Data Collection and Analysis

Data were collected from online reviews about Angkor Wat from TripAdvisor. This is the number one travel review site with 859 million reviews of millions of lodgings, restaurants, cruises, or attractions (TripAdvisor.com, 2020c). Online user reviews (N=32,394) were collected for the period between January 2015 - October 2019 with a 'scraping' technique via a data mining software - RapidMiner Studio.

All the textual reviews contained Visit Date, Review Date, Review Stars, Location ID, Property ID, Review Author, Review Author ID, Review Total (i.e., total number of reviews a user has written on TripAdvisor), Review Author Address, Review Title, and Review Text. R studio and RapidMiner were used to conduct the sentiment and topic analysis. Both of these software tools have been used in previous similar studies (Menner et al., 2016; Pan, 2019; Schmunk et al., 2014). The data analysis process is illustrated in Figure 1.

(Insert Figure 1 Here)

In the sentiment analysis process, ‘retrieve operator’ was utilized to retrieve the textual review data of Angkor Wat, which has also stored information and loaded it for further use. Then, ‘extract sentiment operator’ created sentiment scores using open-source sentiment dictionaries i.e. proprietary API method on existing text attribute. Among four available model for sentiment classification (Aylien, Meaning Cloud, SentiWordNet, and Vader), Vader was used for getting the best results and generalize favorably across context than any other benchmarks (Hutto & Gilbert, 2014). It produced scores based on the dictionary of words utilizing Vader (Valence Aware Dictionary and Sentiment Reasoner) lexicon assigned rule-based sentiments to the score of the text. Vader model disclosed all the words taking part in the scoring, the sum of positive components, the sum of harmful components, and the number of used and unused tokens (Figure 1). Finally, score was provided to the reviews, which indicated the way negative and positively overall analysis of the text in the reviews (anything below a score of -0.05 was tagged as negative and anything above 0.05 was tagged as positive and anything in between inclusively were tagged as neutral).

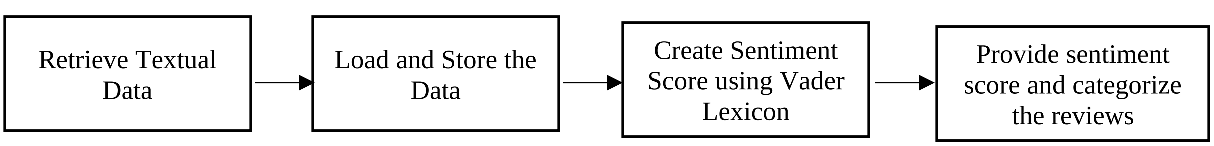

Figure 1. Sentiment Analysis Process

Topic Analysis was firstly conducted on reviews that contained four and five stars and secondly on reviews that contained one, two, and three stars. The ‘read excel’ operator read and load review data from excel. Then, they were filtered for reviews with either four and five stars or one, two and three starts as required. After the ‘select attributes’ operator selected attributes for analysis, the nominal operator converted all nominal attributes to string attributes. Then, drawn existing list of attributes sample was randomly analyzed. The analysis showed that minimum 1000 sample size could generate topics. Then the process document generated word vectors from string attributes. Further, ‘tokenize’ operator splits the text of document into a sequence of tokens based on all non-letter characters, which will result in tokens consisting of one single word. Then ‘filter token (by length)’ was used to filter out those tokens with less than four and more than twenty-five characters. Furthermore, ‘filter stopwords’ removed all tokens equal regular, recurrent but meaningless words (like Angkor, Wat, Siem, Reap, tomb, and raider) from the given file. Then, ‘transform cases’ was utilized to transform the cases of characters to lowercase. Finally, ‘extract data’ extracted topics using Latent Dirichlet Allocation (LDA), the core estimation based on the algorithm of Hoffman. LDA model has been chosen for this process because it allows the inference of topic distribution (Ficamos & Liu, Y., 2016) and also provide presentation of big text data allowing the exploration of different topics and their intensities (Wang, Feng, & Dai, 2018). Since LDA modeling is priori classification method, the number of topics was entered as input following trial and error procedure (Bastani, Namavari,& Shaffer, 2019; Mehta & Vinayagam, 2016). Following the same trail and error mechanism, the topic number 10 was chosen since it provided more meaningful topics accordingly (Figure 2). As for the topic naming, approach used was to manually summarize the meaning of extracted top words. (Maskeri, Sarkar & Heafield, 2008; Shi, Lee & Whinston, 2016).

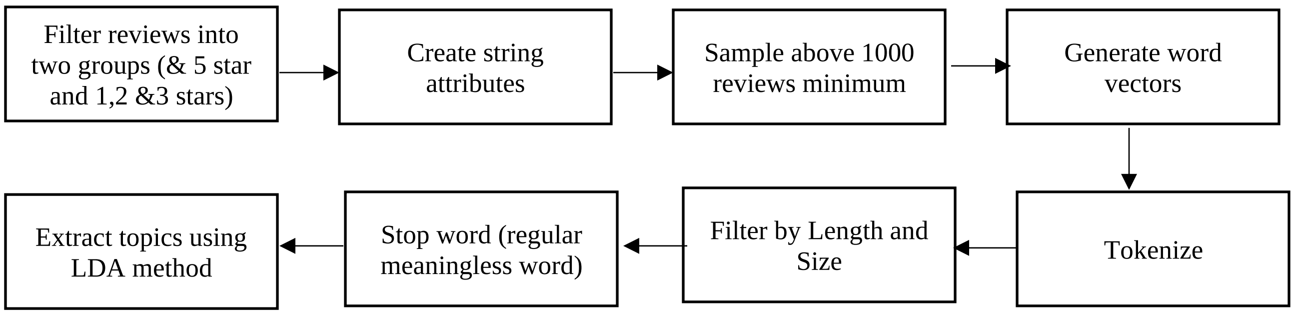

Figure 2. Topic Analysis Process

Results

Sentiment Analysis

Based on the compiled text reviews, 68 reviewers provided a one-star rating, 122 noted as two-star, 871 offered three-star rating, 4,757 reviewers gave four-stars, and 26,576 reviewers noted five-star rating (see Table 1). Furthermore, 1,061 reviews were categorized as negative rating group with one, two, and three-star rating, while 31,333 reviews were grouped as positive rating group with four and five-star rating. These results demonstrate the points were generally positive, while an overwhelming majority rated the attraction as a five-star experience.

(Insert Table 1 Here)

Using the VADER lexicon, the analysis showed that the highest negative score was 17.67, while the highest positive score was 78.95. Also, the highest total score was 61.97, while the lowest total score was -3.897. The total number of positive points received within the entire data set was 70,652.08, and the total negative points was 12,264.95 (see Table 2).

(Insert Table 2 Here)

Further, the Pearson correlation coefficient between the number of stars and total sentiment scores (calculated by subtracting the negative sentiment score from the positive sentiment score) was found to be a weak positive score (0.14). This indicated that irrespective of ratings, most had positive sentiments about the attraction after their visit. Essentially, even if the tourist believed the attraction might have some negative attributes (e.g., crowded), they still left the site with an overall positive sentiment. Result showed that 82 percent gave the attraction five stars, and were 5.7 times more positive than negative as illustrated by the sentiment analysis. Additionally, when the mean total score for each star rating group were calculated, the total means increased with the number of stars (see Table 3).

(Insert Table 3 Here)

Furthermore, ANOVA analysis was conducted to examine whether the star rating groups of reviews were significantly different from each other. Results showed significant differences in the total score mean between groups ‘one and five’, ‘one and four’, ‘three and five’, ‘two and five’, ‘four and five’, ‘three and four’, ‘two and four’, and ‘two and three’ (see Table 4). Basically, there were significant differences between eight of the ten groups in total with mean increases with star ratings, despite a weak correlation between the star ratings and the total sentiment scores. This means that the star ratings for Angkor Wat were effective and corresponded to their sentiments towards the destination. The sentiment analysis obtained from the review corresponded with the star ratings. The overwhelming positive sentiments of visitors towards Angkor Wat is consistent with its selection by TripAdvisor as the number one landmark.

(Insert Table 4 Here)

Topic Analysis

Reviews were categorized as those that provided positive (rating star 4 and 5) and negative (rating 1, 2 and 3 stars) to capture the topics. This priori segment was used to divide the market based on available data (see Dolnicar, 2004), which was self-evaluation of their experiences at Angkor Wat. Further, the star ratings are singled out as fundamental elements for the information evaluation process (Park & Nicolau, 2015). The extracted comments from the reviews along with distinctive words, expression sentiments, and topic names are denoted in Table 5.

(Insert Table 5 Here)

A process to name was employed to derive meaningful topics based on the extracted words. To illustrate, for Topic 1, the words 'sunrise', 'time,' and 'experience' were focused on the meaning of sunrise experience. Likewise, 'time' and 'people' represented surroundings, while the word 'good' imparted positive sentiments. Based on this, Topic 1 was given the name 'Sunrise Experience' with positive sentiments. Therefore, the extracted words, distinctive words, and the type of expressions formed the sources to formulate the topics.

Collectively, ten topics were summarized for those who rated their experience as 1, 2 or 3 stars were – 'sunrise experience', 'guided tours', 'attraction structure, 'clothing style', 'religion origin', 'temple experience, 'world heritage sites', 'crowding and corruptions', 'expenses' and 'persistent selling'. All identified topics were data driven and based on the choice of words used by the reviewers. They were categorized into positive (i.e., sunrise experience, attraction structure, religion origin, and temple experience), neutral (i.e., guided tours, clothing style, world heritage site), and negative expressions (i.e., crowding and corruption, expense and persistent selling). The topics were named based on a minimum of five distinctive meaningful words (see Ren & Hong, 2017).

Similar procedure was followed for those reviewers who rated Angkor Wat with 4 and 5 star ratings. The topics identified were – 'travelling', 'religious origin', 'service and pastime', 'world heritage site' 'clothing style and walking', 'directional', 'guided tours', 'persistent selling’, ‘travel arrangement’ and ‘sunset experience’. Among them, there were positive expressions (e.g., world heritage site, guided tours and sunset experience), neutral (e.g., travelling, religious origin, ‘service and pastime’, ‘directional’ and ‘travelling arrangements) and negative (e.g., persistent selling, clothing style and walking and sunset experience). All were categorized based on the words obtained from extraction. Table 6 illustrates extracted and distinctive words categorized under the topics.

(Insert Table 6 Here)

The results displayed the topics that visitors to Angkor Wat expressed. The priori segmentation of visitors into two broad segments aided to understand the experiences of both satisfied and unsatisfied visitors. Results also revealed both differences and similarities between two types of visitors (i.e., rating 1,2 or 3 vs. rating 4 or 5) with regards to their topic of interest. Satisfied visitors positively associated Angkor Wat as a World Heritage Site, guided tours, and sunset experiences. Conversely, unsatisfied visitors positively associated the site with sunrise experience, attraction structure, temple experience, and religious experience. These findings illustrate that the two segments of visitors had distinctly different experiences. However, both segments had negative experiences due to persistent selling. Moreover, unsatisfied visitors had a negative experience due to crowding and expenses, whereas satisfied visitors had negative experience only due to crowding issues especially during sunrise and sunset. Accordingly, the topic analysis was able to provide insights about visitors’ experiences, and also provides evidence about issues that two segments of visitors faced at Angkor Wat.

Discussion

This study explored visitors’ sentiments and topics registered in TripAdvisor based on their experience at Angkor Wat in Cambodia. The findings showed that an overwhelming majority of the reviews (96%) rated Angkor Wat with four and five stars. Consistently, 85% of the total emotions were positive. It is logical that online reviews were generally highly skewed towards the positive or negative ratings since only extremely satisfied or extremely dissatisfied customers tend to publicize their opinions (Racherla, Connolly, & Christodoulidou, 2013). However, it was interesting to note that reviewers posted reviews with positive sentiments even when they were not satisfied with specific attributes. This could be explained by a previous study which identified lack of high correlation between the overall rating and individual attributes' rating (Racherla et al., 2013). Therefore, tourism practitioners should examine overall rating along with the valence of the reviews on specific attributes to comprehend authentic consumer experience.

Despite a weak correlation between the star rating and the total sentiment scores, significant differences were found between the total mean scores of various star rating groups, except for groups 'one and two' and 'one and three.' This finding validates the approach to separate the data set - one, two, and three stars rating reviews as one group, and four and five-star ratings as the other group. Typically, tourism businesses treat one, two, and three stars rating reviews as grievances, and utilize them to identify reasons for visitor dissatisfaction. Similarly, the rating of four and five should be discerned as the positive performance could contribute to visitor satisfaction. This practice will assist marketers to develop a clear perspective about visitors' sentiments and experiences.

Sunset and sunrise experiences were positive expressions extracted from reviews across all star ratings. This can be attributed to the fact that the architects and engineers carefully designed Angkor Wat around the sun's movement, so that the solar alignments between the temple and a nearby mountaintop shrine provide memorable sunset and sunrise (Sparavigna, 2016). Findings also revealed that visitors mentioned sunset and sunrise experiences with words such as, people, tourists, time, and crowd. This aligns with the traffic statistics recorded at the main gate of the park. Peak periods of traffic at the entrance and exit of the site were associated with the increase in visitor movement during sunrise and sunset (Fletcher et al., 2007). Further, visitors that offered a higher rating also mentioned the word “worth” when they were referred to sunset experiences. It can be concluded that most visitors found sunset or sunrise as a worthy experience despite all other factors.

Temple experience and attraction structure were other positive expressions mentioned in the lower star rating reviews. The distinctive words associated with these two topics reflected the site’s monumental and archaeological remains, temples, characteristic structures, and carved stones as good examples of Khmer architecture. Visitors emphasized the architectural value of Angkor Wat as one of the most important cultural heritage landscapes in Southeast Asia. The term “worth” stood as distinct for topics related to attraction structure. This showed that visitors predominantly valued the attraction structure of Angkor Wat, irrespective of the negative attributes or experiences.

Another topic ‘religious origin’ was voiced as positive expressions in reviews with lower star ratings and as neutral expressions in reviews with higher star ratings. For those with lower ratings, Angkor Wat served as a bucket list item due to its religious value, whereas other visitors gave it a higher overall rating mainly based on its antiquity aspects. Although Angkor Wat is a religious monument, individuals’ visitation patterns influence the meaning assigned to the site, and is subsequently modified by tourism activity over time (Murray & Graham, 1997). This is also evident in other religious sites that have become a destination for visitors (Nyaupane, Timothy, & Poudel, 2015). Basically, due to popularity and/or after inscription as a World

Heritage Site or other designations, religious sites evolve from places that are visited by only adherents to become destinations that are also visited by general tourists. Consistently, common distinctive words such as “history,” “heritage,” and “ancient” were extracted from both positive and negative rating reviews which was appropriately labeled under the topic ‘World Heritage Site’.

The topic “guided tours” were evaluated differently with higher and lower ratings. In lower ratings, the distinctive words included “information,” “history,” and did not reflect any positive emotions. However, for four and five-star ratings, guided tours involved positive expression. According to Weiler & Walker (2014), the essential features of guided tours are to deliver knowledge, provide accurate information, good/new experience, and keep tourists happy or make them enjoy the experience. Thus, these aspects are significant to enhance the tourist experience during guided tours. Furthermore, motivations of heritage site visitors differ, and can affect their preferences as some visit as a “must-see” tourist attraction only (Poria, Biran, & Reichel, 2009). Accordingly, Angkor Wat site management should collaborate with tourism stakeholders in the region to ensure the guided tours not only to enrich visitors’ cultural knowledge, but also enhance their emotional experiences.

The findings revealed that strictly enforced ‘dress code’ in Angkor Wat was one of the issues that visitors complained mostly in both positive and negative reviews. Visitors had to wear appropriate attire with long pants that cover the knees, and shirts over shoulders when visiting temples in Angkor Wat complex. Visitors who gave four or five-star ratings complained about the dress code since long skirts prevented them from ease of walking and climbing steps. As Aulet & Vidal (2018) also emphasized that sacred places are visited by individuals who have diversified motives that range from worship to a transcendental experience. As a result of touristification of religious places, conceptual boundaries between tourism and religion have become increasingly diffuse. Visitors that complained about the official codes of conduct that includes appropriate attire might have different motivations to visit Angkor Wat, namely to experience the architectural, international, local, and cultural value. Conversely, religious site managers have to sustain the spiritual atmosphere of the place for committed visitors that visit with the purpose of praying, meditation and contemplation (Olsen, 2006).

Furthermore, APSARA National Authority which is the government agency that manages the site has formulated and extensively promoted Visitor Code of Conduct. This is disclosed in the site's official website, imprinted on the back of the entrance ticket in detail, as well as prominently displayed across the complex (Angkor.com.kh, 2020). However, it is apparent that a gap between tourist expectations and site management measures exists. This indicates the need for additional initiatives aimed to inform visitors about the code of conduct prior as congruent with sustainability and preservation goals of Angkor Wat.

Reviewers who rated Angkor Wat with one, two, or three stars complained mainly about ticket prices, crowding, and persistent selling at the site. The visitors found ticket prices expensive, and noted persistent selling of vendors highly annoying. Despite some negative comments about ticket prices (US $37 per day), it can be argued that this price is fair when taken into consideration the conservation and preservation costs of such a huge complex, in comparison to other most popular landmarks, museums and heritage sites. As a major safety issue, tourist harassment which includes offensive or distracting behaviors from vendors has negative influences on the image of a destination (Baloglu, Henthorne, & Sahin, 2014; McElroy, Tarlow, & Carlisle, 2007), tourists' willingness to purchase (Henthorne, George, & Smith, 2013), behavioral intentions including word-of-mouth and intention to revisit (Kozak, 2007) and overall experience quality (Alazaizeh et al., 2019; Kozak, 2007). Alrawadieh, Alrawadieh, & Kozak (2019) emphasized that persistent selling is an issue that needs to be resolved with broad perspective and will focus to ensure fair distribution of tourism income while simultaneously create adequate job opportunities. Also, it requires a collaboration of all tourism stakeholders which will aim to educate vendors about the negative impacts of their persistent selling behavior.

With regards to crowding issue, it is dependent on the time, day, and month of visitation. However, crowding is an issue for most popular attractions and leads to visitor dissatisfaction (Q. Jin & Pearce, 2011; Sun & Budruk, 2017). In addition, it deteriorates experience quality especially for purposeful heritage tourists that seeks to gain a deep experience at the site (Alazaizeh et al., 2019). Moreover, massive crowds may impact tourists' perception of heritage sites as a must-see destination and extenuate its value as a World Heritage Site (Larson & Poudyal, 2012). Within sustainable marketing framework, demarketing strategies that involves communications that discourage consumers to visit during certain times or in certain places might be useful for heritage sites that experience overcrowding issues (Donohoe, 2012; Weiler et al., 2019). Information-based crowding management strategies such as publicizing the peak periods and hours on the website and creating sensory maps might both mitigate the overcrowding problem and assuage visitor expectations. GIS-based mobile applications can help site managers to analyze spatio-temporal visitor activities and anticipate potential crowding and congestion issues (Kim et al., 2018).

Limitations and Future Research

The study revealed the sentiments and experiences of Angkor Wat visitors based on their TripAdvisor reviews. There are some limitations to note. First, the study is based on user-generated content which has its own potential biases and subjectivity. Future research could analyze reviews from several sources to compare data or analyze reviews on multiple sites. Second, computer programs for text analysis fail to recognize sarcasm, irony, negotiation, jokes, or exaggeration, whereas these sentiments are identified by a human being. Future studies might employ a content or a thematic analysis procedure with a smaller sample size to especially analyze the negative reviews in more detail. Third, the study utilized reviews posted between the years 2015 and 2019 to generate recent topics; however, future research might compare the data with previous years. Moreover, comparable review might reveal different topics and implications for management, as incremental data might change the interpretation of the cultural attractions. In conclusion, the use of big data analytics based on user generated content should not be the sole basis for managerial decisions, but rather another slice of input along with direct contact field-based visitor intercept surveys.

References

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research, 58*(2) 175–191.

Alazaizeh, M. M., Hallo, J. C., Backman, S. J., Norman, W. C., & Vogel, M. A. (2019). Giving voice to heritage tourists: indicators of quality for a sustainable heritage experience at Petra, Jordan. *Journal of Tourism and Cultural Change*, 17(3), 269-284.

Albert, M. T., & Ringbeck, B. (2015). *40 years World Heritage Convention: popularizing the protection of cultural and natural heritage* (Vol. 3). Walter de Gruyter GmbH & Co KG.

Alrawadieh, Z., Alrawadieh, Z., & Kozak, M. (2019). Exploring the impact of tourist harassment on destination image, tourist expenditure, and destination loyalty. *Tourism Management, 7*3, 13-20.

Anandkumar, A., Foster, D. P., Hsu, D. J., Kakade, S. M., & Liu, Y. K. (2012). A spectral algorithm for latent dirichlet allocation. In *Advances in Neural Information Processing Systems* (pp. 917-925).

Angkor.com.kh. (2020). *Angkor Visitor Code of Condu*ct. Retrieved from http://angkor.com.kh/plan-your-visit/visitor-code-of-conduct/

Aulet, S. & Vidal, D. (2018). Tourism and religion: sacred spaces as transmitters of heritage values, Church*. Communication and Culture, 3*(3), 237-259.

Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, *52*(4), 437-452.

Baloglu, S., Henthorne, T. L., & Sahin, S. (2014). Destination image and brand personality of Jamaica: A model of tourist behavior. *Journal of Travel & Tourism Marketing, 31*(8), 1057-1070.

Barreda, A., & Bilgihan, A. (2013). An analysis of user-generated content for hotel experiences. Journal of Hospitality and Tourism Technology, 4(3), 263–280.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *Journal of Hospitality Marketing & Management, 25*(1), 1–24.

Brigitta, H. S. (2011). *World Heritage Angkor and Beyond-Circumstances and Implications of UNESCO Listings in Cambodia*. Universitätsverlag Göttingen.

Carter, R. W., Thok, S., O'Rourke, V., & Pearce, T. (2015). Sustainable tourism and its use as a development strategy in Cambodia: a systematic literature review. *Journal of Sustainable Tourism*, *23*(5), 797-818.

Clarizia, F., Colace, F., De Santo, M., Lombardi, M., & Pascale, F. (2018, October). A sentiment analysis approach for evaluation of events in field of cultural heritage. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 120-127). IEEE.

Cheng Sokhorng (2 January 2019). *Angkor hosts 2.6M visitors*. The Phnom Penh Post.

Chheang, V. (2011). Angkor heritage tourism and tourist perceptions. *Tourismos*, *6*(2), 213-240.

Christodoulides, G., Jevons, C., & Bonhomme, J. (2012). Memo to marketers: Quantitative evidence for change: How user-generated content really affects brands. *Journal of Advertising Research*, 52(1), 53-64.

Chung, N., & Koo, C. (2015). The use of social media in travel information search. *Telematics and Informatics*, *32*(2), 215-229.

Cox, C., Burgess, S., Sellitto, C. and Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing and Management*, 18(8), 743-764.

Crain, S. P., Zhou, K., Yang, S., & Zha, H. (2012). Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond. In C. G. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (p. 223-258). New York, NY: Springer.

Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising*, *8*(2), 16-25.

Del Chiappa, G., Alarcón-Del-Amo, M. D. C., & Lorenzo-Romero, C. (2016). Internet and user-generated content versus high street travel agencies: A latent gold segmentation in the context of Italy. *Journal of Hospitality Marketing & Management*, *25*(2), 197-217.

Dolničar, S. (2004). Beyond "commonsense segmentation": A systematics of segmentation approaches in tourism. *Journal of Travel Research*, 42(3), 244-250.

Donohoe, H. M. (2012). Sustainable heritage tourism marketing and Canada's Rideau Canal world heritage site. *Journal of Sustainable Tourism*, *20*(1), 121-142.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498-506.

Filieri, R., & McLeay, F. (2013). E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research*, *53*(1), 44-57.

Fletcher, R., Evans, D., Pottier, C., & Rachna, C. (2015). Angkor Wat: an introduction. *Antiquity*, 89(348), 1388-1401.

Fletcher, R., Johnson, I., Bruce, E., & Khun-Neay, K. (2007). Living with heritage: Site monitoring and heritage values in Greater Angkor and the Angkor World Heritage Site, Cambodia. *World Archaeology*, 39(3), 385-405.

Garín-Muñoz, T., & Perez-Amaral, T. (2011). Internet usage for travel and tourism: The case of Spain. *Tourism Economics*, 17(5), 1071-1085. https://doi.org/10.5367/te.2011.0080

García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Information Technology & Tourism*, *16*(1), 45-69.

Gan, Q., Ferns, B. H., Yu, Y., & Jin, L. (2017). A text mining and multidimensional sentiment analysis of online restaurant reviews. *Journal of Quality Assurance in Hospitality & Tourism*, *18*(4), 465-492.

Gkritzali, A., Gritzalis, D., & Stavrou, V. (2018). Is Xenios Zeus still alive? Destination image of Athens in the years of recession. *Journal of Travel Research*, *57*(4), 540-554.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467-483.

Hauser-Schäublin, B. (2011). Heritage Site World Heritage Angkor and Beyond. Göttingen Studies in Cultural Property, Volume 2.

Henthorne, T. L., George, B. P., & Smith, W. C. (2013). Risk perception and buying behavior: An examination of some relationships in the context of cruise tourism in Jamaica. *International Journal of Hospitality & Tourism Administration*, *14*(1), 66-86.

Jalilvand, M. R., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice. *Internet Research*, *22*(5), 591-612.

Jin, S. V., & Phua, J. (2016). Making reservations online: The impact of consumer-written and system-aggregated user-generated content (UGC) in travel booking websites on consumers' behavioral intentions. *Journal of travel & tourism marketing*, *33*(1), 101-117.

Jin, Q., & Pearce, P. (2011). Tourist perception of crowding and management approaches at tourism sites in Xi'an. *Asia Pacific Journal of Tourism Research*, *16*(3), 325-338.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

Kemp, S. (2019). *Digital 2019: Essential insights into how people around the world use the internet, mobile devices, social media, and E-commerce*. Kepios Pte Ltd., We Are Social Ltd. And Hootsuite Inc. https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates

Kennedy, H. (2012) Perspectives on Sentiment Analysis. *Journal of Broadcasting & Electronic Media*, *56*(4), 435-450.

Kim, J., Thapa, B., Jang, S. & Yang, E. (2018). Seasonal Spatial Activity Patterns of Visitors with a Mobile Exercise Application at Seoraksan National Park, South Korea. *Sustainability*, *10*, 2263.

Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management, 72*, 400–410.

Kozak, M. (2007). Tourist harassment: A marketing perspective. *Annals of Tourism Research, 34*(2), 384-399.

Larson, L. R. & Poudyal, N. C. (2012). Developing sustainable tourism through adaptive resource management: a case study of Machu Picchu, Peru. *Journal of Sustainable Tourism*, *20*:7, 917-938.

Lee, M., Lee, S. A., & Koh, Y. (2019). Multisensory experience for enhancing hotel guest experience. *International Journal of Contemporary Hospitality Management*, *31(11),* 4313-4337.

Leung, D., Law, R., van Hoof, H. and Buhalis, D. (2013). Social media in tourism and hospitality: a literature review. *Journal of Travel and Tourism Marketing, 30*(1/2), 3-22.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301-323.

Lim, Y., Chung, Y., & Weaver, P. A. (2012). The impact of social media on destination branding: Consumer-generated videos versus destination marketer-generated videos. *Journal of Vacation Marketing*, *18*(3), 197-206.

Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*(3), 162-172.

McElroy, J. L., Tarlow, P., & Carlisle, K. (2007). Tourist harassment: Review of the literature and destination responses. *International Journal of Culture, Tourism and Hospitality Research*, *1*(4), 305-314.

Mendes-Filho, L., Mills, A. M., Tan, F. B., & Milne, S. (2018). Empowering the traveler: An examination of the impact of user-generated content on travel planning. *Journal of Travel & Tourism Marketing*, *35*(4), 425-436.

Menner, T., Höpken, W., Fuchs, M., & Lexhagen, M. (2016). Topic Detection: Identifying Relevant Topics in Tourism Reviews. In A. Inversini & R. Schegg (Eds.) *Information and Communication Technologies in Tourism* (p. 411–423). Springer, Cham.

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism management*, *43*, 46-54.

Murray, M., & Graham, B. (1997). Exploring the dialectics of route-based tourism: the Camino de Santiago. *Tourism Management, 18*(8), 513-524.

Nyaupane, G. P., Timothy, D. J., & Poudel, S. (2015). Understanding tourists in religious destinations: A social distance perspective. *Tourism Management*, 48, 343-353.

OECD (2007). *Participative web and user-created content: Web2.0, wikis, and social networking*. Paris: Organisation for Economic Co-operation and Development

Olsen, D. G. (2006). Management issues for religious heritage attractions. In D. J. Timothy & D. H. Olsen (Eds.), *Tourism, Religion, & Spiritual Journeys* (p.104-118). New York, NY: Routledge.

Pan, S. (2019). Tourism slogans – Towards a conceptual framework. *Tourism Management*, *72*, 180-191.

Pan, W., Zhong, E., & Yang, Q. (2012). Transfer Learning for Text Mining. In C. G. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (p. 223-258). New York, NY: Springer.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67-83.

Pearce, P. L., & Wu, M. Y. (2015). Soft infrastructure at tourism sites: identifying key issues for Asian tourism from case studies. *Tourism Recreation Research*, 40(1), 120-132.

Poria, Y., Biran, A., & Reichel, A. (2009). Visitors' preferences for interpretation at heritage sites. *Journal of Travel Research*, 48(1), 92-105.

Racherla, P., Connolly, D. J., & Christodoulidou, N. (2013). What determines consumers' ratings of service providers? An exploratory study of online traveler reviews. *Journal of Hospitality Marketing & Management*, 22(2), 135-161.

Ren, G., & Hong, T. (2017). Investigating online destination images using a topic-based sentiment analysis approach. *Sustainability*, *9*(10), 1765.

Schmunk, S., Höpken, W., Fucks, M. & Lexhagen, M. (2014). Sentiment Analysis: Extracting Decision-Relevant Knowledge from UGC. Z. Xiang and I. Tussyadiah (Eds.), *Information and Communication Technologies in Tourism*, (p. 253-265). Springer.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, *32*(5), 608-621.

Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017). Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*, 8(2), 220-238.

Song, H. & Liu, H. (2017). Predicting Tourist Demand Using Big Data. In Z. Xiang, D.R. Fesenmaier (Eds.), *Analytics in Smart Tourism Design: Concepts and Methods* (p. 13-30). Switzerland: Springer International Publishing.

Sparavigna, A. C. (2016). Solar Alignments of the Planning of Angkor Wat Temple Complex. *Philica*, Article, (591).

Sun, Y. Y., & Budruk, M. (2017). The moderating effect of nationality on crowding perception, its antecedents, and coping behaviours: A study of an urban heritage site in Taiwan. *Current Issues in Tourism*, *20*(12), 1246-1264.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*, 550-568.

TripAdvisor.com (2020a*). TripAdvisor Has a Massive Influence on a $5 Trillion+ (and Growing) Global Travel Economy, According to New Oxford Economics Study*. Retrieved from http://ir.tripadvisor.com/news-releases/news-release-details/tripadvisor-has-massive-influence-5-trillion-and-growing-global

TripAdvisor.com (2020b). *10 Best World Landmarks To Add To Your Travel Bucket List—And How To See Them*. Retrieved from https://www.tripadvisor.com/blog/best-world-landmarks-famous-attractions-travel-bucket-list/

TripAdvisor.com (2020c). *About Tripadvisor*. Retrieved from https://tripadvisor.mediaroom.com/IN-about-us

UNWTO (2020). Tourism and culture. Retrieved from https://www.unwto.org/tourism-and-culture

Van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, culture & society*, *31*(1), 41-58.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51-65.

Xiang, Z., Wang, D., O'Leary, J. T., & Fesenmaier, D. R. (2015). Adapting to the internet: trends in travelers' use of the web for trip planning. *Journal of Travel Research*, *54*(4), 511-527.

Valdivia, A., Martínez-Cámara, E., Chaturvedi, I., Luzón, M. V., Cambria, E., Ong, Y. S., & Herrera, F. (2020). What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules. *Journal of Ambient Intelligence and Humanized Computing*, *11*(1), 39-52.

Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, *29*, 142-156.

Weiler, B., Moyle, B. D., Scherrer, P., & Hill, M. (2019). Demarketing an iconic national park experience: Receptiveness of past, current and potential visitors to selected strategies. *Journal of Outdoor Recreation and Tourism*, *25*, 122-131.

Weiler, B., & Walker, K. (2014). Enhancing the visitor experience: Reconceptualising the tour guide's communicative role. *Journal of Hospitality and Tourism Management*, *21*, 90-99.

Winter, T. (2007). Rethinking tourism in Asia. *Annals of Tourism Research*, *34*(1), 27-44.

Worldwide Tourism Economics. (2018). Sizing Worldwide Tourism Spending (or "GTP") & TripAdvisor's Economic Impact. Retrieved from https://mk0tainsights9mcv7wv.kinstacdn.com/wp-content/uploads/2018/09/Worldwide-Tourism-Economics-2017-compressed.pdf

Wu, M. Y., & Pearce, P. L. (2016). Tourism blogging motivations: Why do Chinese tourists create little "Lonely Planets"? *Journal of Travel Research*, *55*(4), 537-549.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human behavior*, *27*(2), 634-639.

Yoo, K. H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior*, *27*(2), 609-621.