e-Review of Tourism Research (eRTR), Vol. 16, No. 4, 2019
http://ertr.tamu.edu

Sina Shokoohyar
Saint Joseph's University

Determinants of Backpackers' Perceptions of Security? A WOM-based Approach

This paper aims to identify the determinants of backpackers' perceptions of security. In particular, the impact of hostel cleanliness, location, staff, and atmosphere as well as backpackers' country of origin on backpackers' perceptions of security are being studied. Using Word-of-mouth (WOM) approach, the analysis shows that hostel cleanliness, location, staff, and atmosphere have a positive and significant impact on backpackers' perceptions of security. Additionally, backpackers from countries with a high level of security have higher expectations about hostel security and in turn are harder to satisfy compared to backpackers from countries with lower security. Applying Centering Resonance Analysis (CRA) and Naïve Bayes classification, this study provides evidence on how these attributes impact backpackers' perceptions of security. Results of this study are valuable in demonstrating how hostel managers can utilize rich information generated by backpackers on the Internet to develop and improve their business.

Key words: Hostels, Security, Backpackers, Country of Origin, Word of Mouth, Centering Resonance Analysis, Sentiment Analysis.

Sina Shokoohyar
Erivan K. Haub School of Business
Saint Joseph's University
Philadelphia, PA 19131
United States
Phone: [001] (610) 660 2217
Email: sshokooh@sju.edu

Sina Shokoohyar is an Assistant Professor of Decision and System Sciences (DSS) at Saint Joseph's University. His research focuses on developing decision analytic tools that can be implemented in ride-sharing platforms (e.g., Uber, Lyft, …), tourism industry, new product development project, and supplier evaluation systems.

Introduction

Youth and student market accounts for 20 to 25 percent of international tourism as a whole (Hecht & Martin, 2006; Richards, King, & others, 2003). Hostels provide backpackers with budget accommodation services combined with an informal atmosphere. Online booking systems are a key factor in connecting hostels around the world and keeping their cost low and their systems efficient. Hostelworld.com, a leading online booking system maintains over 14,000 hostels around the world. This reservation system generated revenue of €21.5 million in 2016 (Hostelworld, 2016).

It is important for hostels to identify the factors impacting backpackers' choices of hostels to expand their business. Several studies have shown the importance of security in the backpackers' choice of a hostel (Amblee, 2015; Cró & Martins, 2017; Cró, Martins, Simões, & Calisto, 2018; Shanahan & Hyman, 2007). Backpackers may analyze a hostel's security by going over previous guest reviews available on online booking systems or may simply focus on the average security rate provided. Research has shown that in this network of hostels, World-of-mouth (WOM) recommendations significantly impact the backpackers' choice of a hostel (Martins, Rachão, & Costa, 2018). Hostelworld.com provide over 9 million post-stay reviews since 2005.

Analyzing WOM recommendations are important for backpackers, hostel managers, and online booking systems: First, form backpackers point of view, the average rating along with the review content shows what they can expect to experience in a given hostel. Second, from hostels managers' point of view, analyzing the review contents help them to explore reasons behind low/high ratings and in turn to identify their hostels' strengths and weaknesses, and prescribe appropriate corrective strategies. Through analyzing interview contents, Hecht and Martin (2006) showed that lockers, safety deposit boxes, front desk (24 hours), and locks on doors are the main satisfaction factors of safety and security at hostels. Improving these items (i.e., lockers, safety deposit boxes, front desk (24 hours), and locks on doors) by hostels' managers will help hostels to receive a higher rating in security and in turn may improve guests' willingness to pay (Cró et al., 2018). Finally, online booking systems may integrate a summary of the review content on their portal along with the average rating to facilitate backpackers' hostel picking process.

Security is defined as threats imposed by people, and that is why security is a dynamic element, meaning that it depends on the person posing a security threat, and it cannot often be foreseen. Backpackers' perception of security is determined by how safe and secure they feel about themselves and their belongings. Guest expectations are defined as beliefs about service delivery that serve as standards or reference points against which performance is judged (Wilson, Zeithaml, Bitner, & Gremler, 2008), and customer perceptions are subjective assessments of actual service experiments through interaction with the others (Wilson et al., 2008). Amblee (2015) showed that hostels cleanliness and location positively and significantly impact backpackers' perception of security. Clean facility implies control and organization, and in turn improves backpackers' perception of security. It is also shown that the location of a hotel (safe/unsafe neighbourhood) trigger thoughts regarding personal security (Shortt & Ruys, 1994). Some evidence shows that there might be other factors that can also impact backpackers' perceptions of security. A guest left this review on Hostelworld.com: "Original staff stole all the money whilst there. Some of the staff were very rude." This review shows how important a hostel's staff is in making a guest feel insecure. Another guest left a review on Hostelworld.com about hostel atmosphere: "This place was not as described on the internet and it was very smelly and the people staying here appeared to be very "unsavoury" I would not have felt safe here and left as soon as I saw the place.". This review indicates how an unfriendly atmosphere can impact backpackers' perceptions of security. Staff, guests, and outside visitors all pose potential security issues for hostels.

Motivated by the above observations, this study aims to investigate the following research questions regarding determinants of backpackers' perceptions of security. First, how do hostel cleanliness, location, staff and atmosphere impact backpackers' perceptions of security? Second, are backpackers' different in their expectations of security?

To address these research questions, 324,321 guests' reviews from Hostelworlds.com are collected. Guests' country of origin and hostel country are categorized to safe and unsafe countries based on the security rating provided by travel.state.gov. The results show that hostels' cleanliness, location, staff, and atmosphere as well as backpackers' country of origin significantly impact backpackers' perceptions of security. Better hostel cleanliness, location, staff, and atmosphere can make guests feel more secure. Additionally, a guest who is coming from safe countries have a higher expectation of security and in turn give a low rating to hostels in term of security. Using Centering Resonance Analysis (CRA) and Naïve Bayes classifier this study provides evidence of how these factors impact backpackers' perceptions of security.

The empirical results suggest that managers can improve their hostel in other factors such as cleanliness, location, staff and atmosphere and in turn make their guests feel more secure. Depending on the targeted market, guests' country of origin can help managers to plan and target a guest's level of desired security.

Literature Review

Several studies identified security as an important attribute that impacts backpackers' opinions while choosing a hostel. Cró et al. (2018) studied the impact of security on backpackers' willingness to pay. They observed that backpackers are willing to pay more in the least peaceful countries. Additionally, women and older guests have a higher willingness to pay. Cró and Martins (2017) explored the impact of guest reviews on the hostel's price premiums. They observed that backpackers are willing to pay more for security when the hostel is located in European countries with higher crime indexes. Amblee (2015) studied the impact of cleanliness and location on backpackers' perceptions of security. He showed that both cleanliness and location positively and significantly impact backpackers' perception of security. Additionally, he observed a small country effect. In his study, he included hostels that are located in South Korea, Thailand, Cambodia, UK, and Hong Kong. Shanahan and Hyman (2007) identified the important attributes that impact American tourists' overall satisfaction when traveling to Ireland and China. They observed that Americans would offset their expectations about cleanliness and price for an increase in security when traveling overseas.

This study differs from the aforementioned research papers in two main points. First, this study not only analyzes the impact of hostel cleanliness and location (like previous studies) but also explores the impact of other factors like hostels' staff and atmosphere on backpackers' perceptions of security. Second, by applying sentiment analysis on guest review content, this study provides evidence on how hostel staff impacts backpackers' perceptions of security.

Some studies also show that backpackers should not be treated as a homogenous group. Dayour et al. (2016) explored the determinant of backpackers' expenditure. They identified nationality, culture and demographic characteristics as significant determinants of backpackers' expenditure. Therefore, they conclude that backpackers should not be treated as homogenous. Oliveira-Brochado and Gameiro (2013) studied the impact of backpackers' age, gender and nationality on their travel motivations. Their findings show the existence of increasing heterogeneity among backpackers' preferences. Hecht and Martin (2006) studied the impact of backpackers' demographic characteristics on their service preferences when they travel to Canada. They observe that young backpackers (15-25 years old) consider backpacking as a more social and cultural experience compared to older backpackers. Additionally, they showed that older backpackers are more concerned about privacy and in turn are willing to pay more for privacy. They also observe that Australians and Europeans expect less hotel-typical services than North/South Americans and Asians. Examples of literature studying the different aspect of backpackers' diversity include Social identity of Chinese backpackers (Zhang, Morrison, Tucker, & Wu, 2018), identity construction of backpackers (Zhang, Tucker, Morrison, & Wu, 2017), impact of backpackers' age and hostel service quality on guests' satisfaction (Lima, Vicente, & others, 2017), literature review on Asian female travelers (Yang, Khoo-Lattimore, & Arcodia, 2017), heterogeneity among backpackers (Brochado & Rita, 2016), role of culture and nationality (Maoz, 2007), racialized and gendered nature of backpacking (Teo & Leong, 2006), and looking local (Muzaini, 2006).

This study considers both the hostel's country effect as well as backpackers' country of origin effect. Considering both attributes facilitates studying of the interaction effect. Therefore, backpackers' expectation of security based on their country of origin is studied.

Data collection and methodology

Data collection

Reviewers' data is collected from Hostelworld.com. Through this website, backpackers can book hostels and leave reviews regarding their experience of their stay at the hostel. Backpackers can rate hostels in 7 dimensions: overall rating, value for money, security, location, facility, staff, and atmosphere. These ratings can be anything in a range of 0 to 10 stars. In this study, reviewer rating in all seven dimensions as well as the reviewer's country of origin and the review content are collected. In total 324,321 reviews are collected from 24,658 hostels. Note that the overall rating is just a linear combination of other ratings, and therefore it is not included in any of the analysis. After cleaning the data (Excluding reviews with not specifically and correctly mentioned country of origin), 321,366 reviews are left in the dataset. The summary of the extracted attributes is presented in Table 1.

To categorize the countries based on their level of safety and security, the four-level security rating provided by travel.state.gov is used. This dataset categorizes countries based on their level of safety and security to travel into four levels: 1: exercise normal precautions, 2: exercise increased caution, 3: reconsider travel and 4: do not travel. The dataset is available at

Travel.State.Gov. Countries are organized into two levels: safe countries with the safety levels of 1 and 2, and unsafe countries with the safety levels of 3 and 4.

Figure 1 shows safe countries (in black) and unsafe countries (in gray). Table 2 represents the distribution of guests and hostels over the safe and unsafe countries.

Table 1. Extracted Attributes

Attribute	Description
Reviewer ID	An ID that is assigned to the guest by Hostelworld.com
Country of Origin	Reviewers' country of origin
Hostel Name	Name of the hostel
Hostel Country	Country that the hostel is located in
Rating Dimensions	Overall Rating, Value for Money, Security, Location, Facility, Staff, Atmosphere
Review Content	Review content written by the reviewer

Figure 1. Safe Countries for Traveling (Black Area) and Unsafe Countries to Travel (Gray Area)

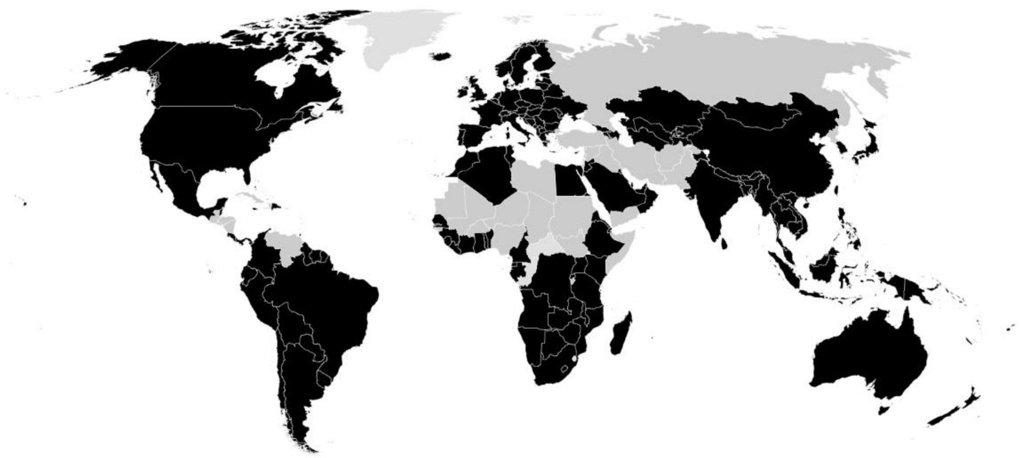


Table 2. Distribution of Backpackers' Country of Origin and Hostel Locations

		Hostel Location	
		Safe Country ■	**Unsafe Country ■**
Reviewer Country of Origin	Safe country	312827	6176
	Unsafe Country	2243	120

Methodology

To study the impact of hostel location, cleanliness, staff, atmosphere, facility, and value for money as well as backpackers' country of origin on their perceptions of security, regression analysis is used. As facility and value for money are highly correlated with VIF of 28.46, a new variable facility×value for money is created and used as a control variable. To control for hostels' effect, hostels' fixed effect is incorporated. Model's specification is as follows:

$$\begin{aligned} &Security \\ &= \beta_0 + \beta_1[Location] + \beta_2[Cleanliness] + \beta_3[Atmosphere] + \beta_4[Staff] \\ &+ \beta_5[Facility \times Value\ for\ Money] + \beta_6[1\ if\ Reviewer\ is\ Coming\ from\ a\ Safe\ Country] \\ &+ \beta_7[1\ if\ Hostel\ is\ in\ a\ Safe\ Country] \\ &+ \beta_8[1\ if\ Reviewer\ is\ Coming\ from\ a\ Safe\ Country\ and\ Hostel\ is\ in\ a\ safe\ country\ ] \\ &+ \epsilon \end{aligned} \tag{1}$$

Centering resonance analysis (CRA) (Corman, Kuhn, McPhee, & Dooley, 2002) is applied to identify important words in guests' reviews and link these words into a network. CRA is a network text analysis that consists of three steps: selection, linking, and indexing. In the selection step, instead of looking at the whole review content, only noun phrases are selected. A noun phrase is a noun and additional words in the sentence that modify it. Words that can modify a noun include nouns, adjectives, and determiners (i.e., the, an, a, etc.). In this process, every sentence is converted to one or more noun phrase. In this step, all extracted nouns and adjectives that are part of a noun phrase are considered to be nodes of the network. In the second step, linking, any two words of a noun phrase are linked with an edge. The existence of the edge shows how words are connected in the review content.

In the third step, indexing, all nodes (i.e., words in the noun phrase) are indexed based on the betweenness centrality. Betweenness centrality of node v ($\sum_{u \neq v \neq w} \frac{\sigma_{uw}(v)}{\sigma_{uw}}$) measures the extent to which a node lies on paths between other nodes. σ_{uw} denotes the total number of shortest paths from node u to node w, and $\sigma_{uw}(v)$ denotes the number of those paths that pass through v. To normalize the index, the betweenness centrality is divided by $\frac{(N-1)(N-2)}{2}$ where N is the number of nodes. High betweenness centrality of the word shows that it has been the main concern in backpackers' reviews, and has been repeated in many different noun phrases. Before applying CRA, reviews in other languages except English are removed. In total 171,273 reviews are included in the analysis. In all analysis also stem of a word is considered. The idea of a stemming works as a normalizing method. The idea is that all words with the same stem convey the same meaning and therefore all are converted to their stem.

As an example, consider these two reviews: 'Some staff were stealing from guests.' and 'Excellent staff and safe place to stay.'. In the first review the words staff and guest, and in the second review excellent, staff, safe, and place are the adjectives and nouns. In the selection step of CRA, all nouns and adjectives are considered as a node of the network. In the second step of CRA, linking, an edge is considered between any two words selected from each review. Therefore, the edges are: (staff, guests), (excellent, staff), (excellent, safe), (excellent, place), (staff, safe), (staff, place), and (safe, place) where each (u, w) denote an edge between node u and node w. In the third step, indexing, using betweenness centrality, nodes' indexes of staff, the words guests, excellent, safe and place are calculated as 0.5, 0.0, 0.0, 0.0, and 0.0, respectively. Staff has the highest betweenness centrality and therefore is selected as the most important feature talked about in the two reviews.

To analyze the sentiment of guest reviews, Naive Bayes Classifier is used to organize the message conveyed through the review content. Sentiment analysis is widely used to analyze online reviews (Pang & Lee, 2005; Pang, Lee, & others, 2008; Prabowo & Thelwall, 2009; Shokoohyar, 2018; Ye, Zhang, & Law, 2009; Yu, Liu, Huang, & An, 2012). Naive Bayes classifier is a popular and simple machine learning method for text classification and performs well in many domains (Domingos & Pazzani, 1997; Go, Bhayani, & Huang, 2009). For a literature review on opinion mining and sentiment analysis approaches, the readers are referred to (Liu & Zhang, 2012; Vinodhini & Chandrasekaran, 2012).

In the Naïve Bayes Classification approach, the review is considered as a bag-of-words. Each word in a review are included in the bag-of-words; words are unordered, and their position in the document is ignored. In this method, only the words' frequency is used for analysis. Reviews are labeled based on security rating; Reviews with 0 to 5 stars are labeled as negative reviews, 6 stars are labeled as neutral, and 7 to 10 star reviews are labeled as positive reviews. In the analysis, 90% of the reviews are considered for training, and remaining for testing the classifier. The Naïve Bayes classifier uses the training review set to extract important features that are related to each class and uses these features to predict the class of given reviews from the test set. Then the test set is used to measure the accuracy of the classifier. In the analysis, only positive and negative reviews are used. The set of classes is denoted as $C = \{Neg, Pos\}$ where neg stands for negative reviews and pos stands for positive reviews. The Naïve Bayes classifier returns the class $\hat{c}$ with the highest posterior probability given the review, R, i.e., $\hat{c}(R) = \underset{c \in C}{\operatorname{argmax}} P(c|R) = \underset{c \in C}{\operatorname{argmax}} P(R|c)P(c)$. Note that the last equality follows from the Bayes rule. Without a loss of generality review, R can be presented as a set of features: $w_1, \ldots, w_n$ (word or feature i is denoted as w_i) and therefore $\hat{c}(R) = \underset{c \in C}{\operatorname{argmax}} P(R|c)P(c) = \underset{c \in C}{\operatorname{argmax}} P(w_1, \ldots, w_n|c)P(c)$. Using Naïve Bayes assumption, that is $P(w_i|c)$ are independent given class c, $\hat{c}(R)$ can be simplified as $\hat{c}(R) = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{w \in W} P(w|c)$ by using maximum likelihood estimation with Laplace smoothing $P(c) = \frac{N_c}{N_R}$ and $P(w_i|c) = \frac{1+n(w_i,c)}{|V|+\sum_{w \in V} n(w,c)}$. N_c is denoted as the number of reviews with the class of c in the training set of reviews, N_R as the total number of reviews, $n(w_i, c)$ as the frequency of w_i in class c, and V as the union of word types in all classes.

Results

The summary statistics of the variables used in this study is shown in Table 3. Before proceeding to the regression analysis presents the correlation between any two rating categories.

Table 3. Descriptive Statistics

Rating Category	Count	Mean	Std
Overall Rating	324,321	8.01	1.79
Value for Money	324,321	1.34	3.19
Security	324,321	8.19	2.08
Location	324,321	8.44	2
Facility	324,321	1.29	3.07
Staff	324,321	8.3	2.25
Atmosphere	324,321	7.31	2.38
Cleanliness	324,321	8.02	2.35

Figure 2. Correlation Matrix of Hostels Rating Categories

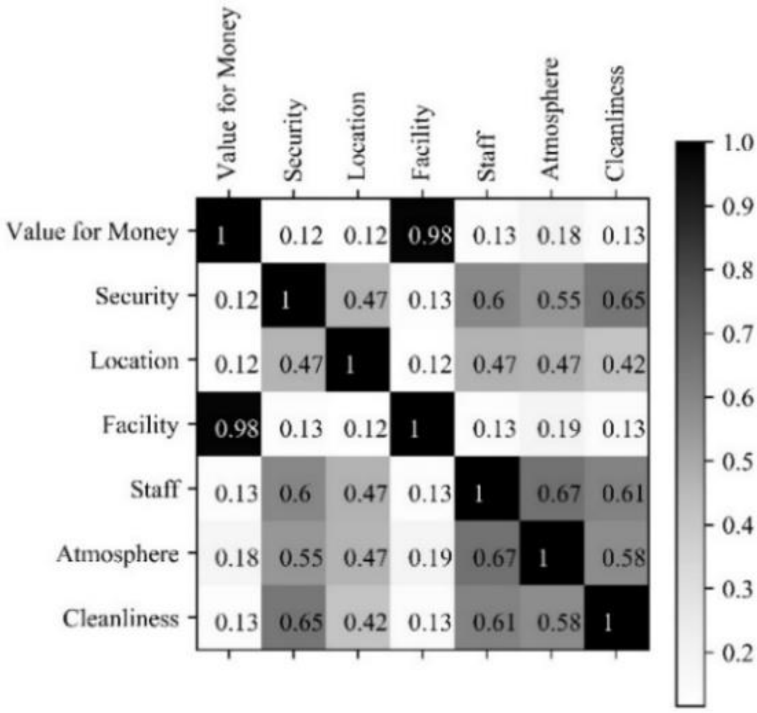


Figure 2 shows that rating categories can be organized into two groups based on the correlation among them. Value for money and facility in one group, and security, location, staff, atmosphere, and cleanliness in the second group. This study focuses only on backpackers' perception of security, and therefore it only includes the second group of variables (i.e., security, location, staff, atmosphere, and cleanliness). Before, moving to the regression analysis, note that the Variation Inflation Factor (VIF) among variables are checked and are all less than 3. Therefore multicollinearity is not very high.

Table 4 represents the summary of the regression analysis. The dependent variable is the hostel security rated by the backpackers, and the independent variables are shown in the first column. The analysis provides two main findings. First, the coefficient of location, cleanliness, atmosphere, staff, and facility×value are positive and significant. This result shows that cleanliness with a coefficient of 0.289 has the strongest impact on security. Location with a coefficient of 0.197 is the second most important determinant of security. This result is in line with (Amblee, 2015). He showed that cleanliness and location both have a significant impact on backpackers' perception of security. Our results further show that atmosphere, staff, and facility×value also have a significant impact on security. Second, the result shows that backpacker's country of origin has a negative and significant impact on their perception of security. This result indicates that backpackers who are coming from safer countries are more difficult to satisfy in terms of security. The result shows that backpackers' origin has less of an impact on security if the hostel is located in a safe country. Findings show that attributes such as location, cleanliness, atmosphere, staff, facility, and value for money can compensate for low security. On the other hand, backpackers' country of origin is out of a hostel's control. The result implies that the backpackers who are coming from safer countries expect higher levels of security, therefore depending on the hostels' targeted market, owners can plan for the desired level of security.

Table 4. Summary of the Regression Analysis

Independent Variables	*β*	Std	p-value
Location	0.197	0.001	0.000
Cleanliness	0.289	0.001	0.000
Atmosphere	0.105	0.001	0.000
Staff	0.185	0.001	0.000
Facility × Value for money	0.000	0.000	0.000
1 if Reviewer is Coming from a Safe Country	-0.362	0.142	0.011
1 if Hostel is in a Safe Country	-0.424	0.256	0.098
1 if Reviewer is Coming from a Safe Country and Hostel is in a Safe Country	0.305	0.145	0.036
Intercept	2.354	0.252	0.000
R-Squared	0.514		
Number of Observation	315,070		

Figure 3. Distribution of Review Content Length

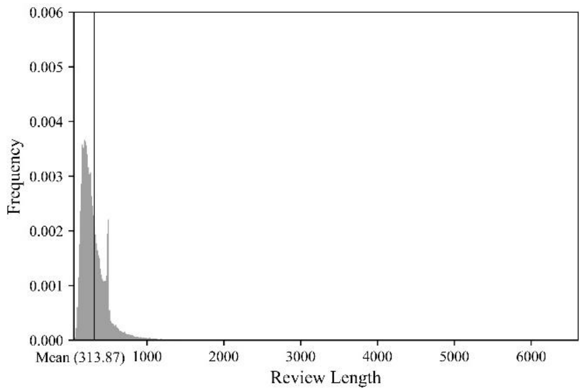
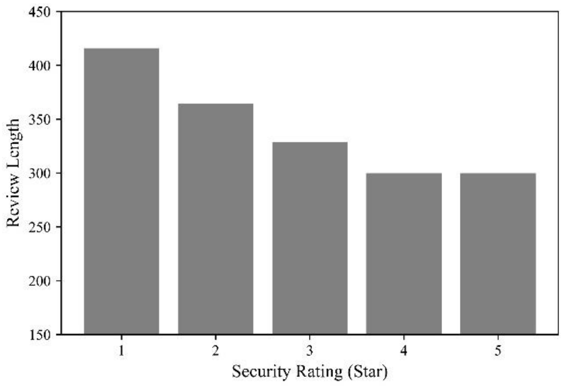


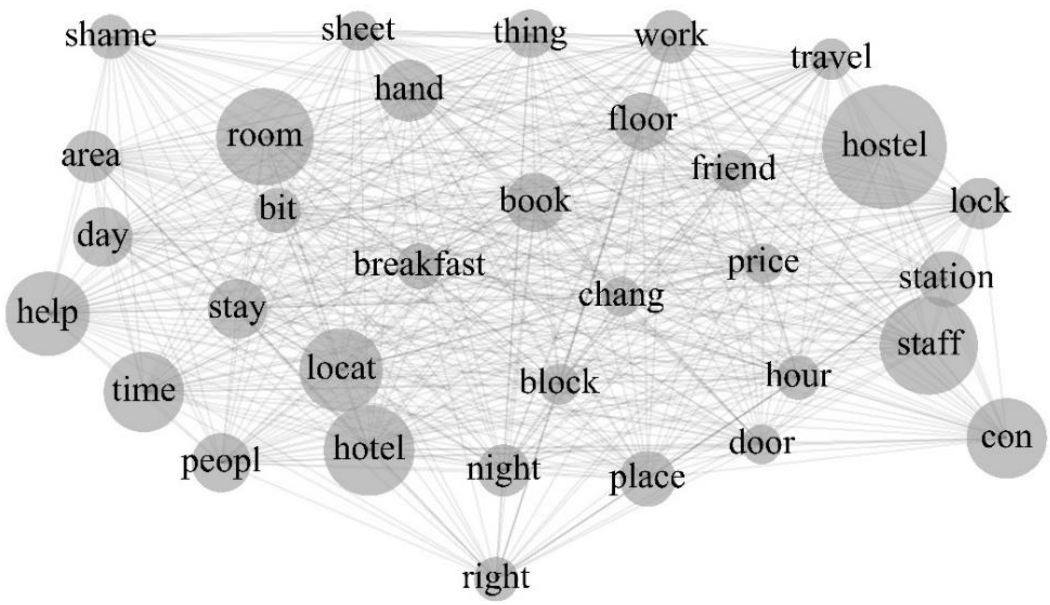
Figure 4. Frequency of Review Content Length on Security Rating



To further study the backpackers' perception of security, backpacker review content is analyzed. In our analysis, the reviews in languages besides English are removed. In total, 171,273 reviews are analyzed. Figure 3 shows the distribution of review content length. The review content length is ranged from 14 words per review to 4349 words per review, with a mean of 313.87 words. Figure 4 represents the average review length of security rating. It shows that backpackers who gave hostels a very low-security rating are more informative and try to elaborate on their concerns by providing longer reviews.

Using betweenness centrality, Figure 5 shows the top commonly used words in reviews with security ratings of 0 to 5. Recall that reviews are organized with security considering ratings of 0 to 5 stars as negative reviews, 6 stars as neutral reviews, and with 7 to 10 stars as positive reviews. The size of each node corresponds to the betweenness centrality of the word, and shows the importance of that word in the network. The result shows a high correlation between backpackers' perception of security and other attributes, such as a hostel's location, cleanliness, atmosphere, and staff. The result of this analysis is summarized in Table 5.

Figure 5. Most Commonly Used Words in Backpackers' Reviews with Negative Reviews



In Table 5, each column shows the correlation between the existence of each word and rating dimensions. Note that the security rating sorted is ascending. The word Lock is negatively correlated with security, and it shows that the existence of this word in the review negatively impacts backpackers' perception of security. On the other hand, the word Help is positively correlated with security. The result shows that a variety of attributes impact backpackers' perception of security. Besides cleanliness (words like: Sheet, Room, Floor, etc.)

and location (words like: Locate, Station, Area, etc.), the results show that the staff and atmosphere of hostels also have a great impact on the backpackers' perceptions of security.

Table 5. Summary of Most Commonly Used Words in Backpackers' Review with Negative Reviews

Word	Betweenness Centrality	Value for Money	Security	Location	Facility	Staff	Atmosphere	Cleanliness
Lock	0.02	0.02	-0.16	-0.01	0.02	-0.06	-0.04	-0.06
Door	0.01	0.00	-0.14	-0.04	0.00	-0.08	-0.08	-0.08
Book	0.03	-0.01	-0.12	-0.09	-0.01	-0.16	-0.12	-0.11
Room	0.07	0.02	-0.12	-0.02	0.02	-0.11	-0.15	-0.15
Night	0.02	0.02	-0.08	-0.06	0.02	-0.08	-0.08	-0.08
Sheet	0.01	-0.02	-0.08	-0.05	-0.02	-0.09	-0.08	-0.14
Floor	0.02	0.01	-0.07	-0.01	0.00	-0.06	-0.08	-0.11
Work	0.02	0.02	-0.06	-0.02	0.02	-0.07	-0.05	-0.07
Hour	0.01	-0.02	-0.06	-0.08	-0.02	-0.09	-0.07	-0.06
Con	0.05	0.02	-0.06	-0.03	0.01	-0.08	-0.08	-0.06
Change	0.01	-0.01	-0.06	-0.04	0.00	-0.08	-0.06	-0.08
People	0.03	0.00	-0.04	0.00	0.00	-0.02	0.01	-0.04
Day	0.03	0.01	-0.03	-0.02	0.01	-0.05	-0.03	-0.04
Thing	0.02	0.00	-0.02	0.01	0.00	-0.02	-0.01	-0.02
Hand	0.03	0.01	-0.02	0.00	0.01	-0.02	-0.02	-0.02
Hotel	0.06	-0.04	-0.01	-0.05	-0.04	-0.04	-0.11	0.00
Price	0.01	-0.01	-0.01	-0.01	-0.01	-0.02	-0.04	-0.02
Shame	0.01	0.00	-0.01	0.00	0.00	-0.02	-0.01	-0.02
Block	0.01	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01
Hostel	0.11	0.06	0.01	0.02	0.06	0.01	0.06	0.03
Time	0.05	0.00	0.01	0.02	0.00	0.00	0.05	0.01

Word	Betweenness Centrality	Value for Money	Security	Location	Facility	Staff	Atmosphere	Cleanliness
Area	0.02	0.07	0.01	-0.01	0.07	0.02	0.01	0.03
Bit	0.02	0.02	0.01	-0.03	0.02	0.01	-0.01	-0.01
Right	0.01	0.00	0.02	0.05	0.00	0.03	0.02	0.03
Travel	0.01	0.02	0.03	0.02	0.02	0.03	0.04	0.04
Place	0.02	0.00	0.04	0.02	0.00	0.05	0.05	0.05
Station	0.02	0.01	0.04	0.00	0.01	0.04	0.00	0.06
Breakfast	0.02	0.07	0.06	0.04	0.07	0.04	0.02	0.06
Stay	0.03	0.03	0.07	0.07	0.03	0.09	0.1	0.08
Locate	0.05	0.04	0.07	0.17	0.04	0.07	0.04	0.06
Staff	0.07	0.06	0.08	0.09	0.06	0.10	0.11	0.09
Friend	0.01	0.05	0.10	0.09	0.05	0.15	0.13	0.12
Help	0.05	0.07	0.14	0.11	0.07	0.19	0.14	0.15

Here are some of the examples of reviews that show how staff and atmosphere damaged backpackers’ perception of security:

Example 1: “DO NOT STAY HERE!I Caught one of the staff member stealing money and shoes from my room.”

Example 2: “The staff steal money from you! they pretend they can’t speak English and they can! they are always aware of what is going on, and they are they always try steal things!”

Example 3: "There were holes in the wall, and they wake you up every day around 9 and take your bed. The shower near my room didn't have a knob on it to control the temp, and most of the toilet seats were not attached to the toilets. Staff also likes to steal your stuff."

Example 4: "I gave security a low grading because I feel that the extra keys of the rooms shouldn't be hanging in the open in the staff office. Anyone could have just gone in, distract the receptionist guy, and steal the keys! Other than that, didn't expect to clear the bed sheets myself too."

Example 5. "This place was not as described on the internet and it was very smelly, and the people staying here appeared to be very "unsavoury" I would not have felt safe here and left as soon as I saw the place."

Example 6. “DO NOT STAY HERE! Most of the people who stay here are long term residents. I had to call the police while staying here as putting it nicely you would say that these residents are not of good character. The police told me that they are frequently called to this hostel either because the long term residents are picking on someone staying there or are fighting amongst themselves. Be warned, you stay here at risk for your safety. Better choices of places to stay would be some of the commercially run hostels such as YHA, St Christophers, The Generator. These may cost a few pounds more, but you will have a much better experience of London."

The above examples show that staff are either accused of stealing backpackers' personal items (examples 1 to 3), or they are accused of not protecting backpackers' belongings (example 4). Examples 5 and 6 show how the atmosphere of the hostel can negatively impact backpackers' perceptions of security. Reviewers of these examples (1 through 6) rated the security of the hostels that they stayed in as 2, 2, 4, 4, 2, and 2 stars, respectively. Note that in the first 4 examples the core of the review is about staff (in Italic), and in examples 5 and 6 the core of the review is about the people.

Using the Naïve Bays Classifier, the security sentiments of reviews are classified in positives and negatives. The result of the classification is presented in Table 6 along with the enumeration of the feature sets in the last column. Note that the Pos and Neg stands for Positive and Negative, respectively. The results show that cleanliness features are among the most important features that determine backpackers' perception of security. 90% of the reviews are included to train the classifier, and the remaining are for testing it. The accuracy of the classifier is 79.22%.

Table 6. 20 Most Informative Features (Extracted Using Naïve Bays Classifier)

Feature	Likelihood	
Disgust	Neg:Pos	88.9:1
Crap	Neg:Pos	77.2:1
Horrible	Neg:Pos	49.9:1
Filthy	Neg:Pos	43.1:1
Worst	Neg:Pos	40.8:1
Shit	Neg:Pos	40.2:1
Unhelpful	Neg:Pos	35.3:1
Angry	Neg:Pos	32:1
Mould	Neg:Pos	32:1
Sucked	Neg:Pos	32:1
Disgusting	Neg:Pos	31.9:1
Sucks	Neg:Pos	30.1:1
Awful	Neg:Pos	29.1:1
Unclean	Neg:Pos	26.7:1
False	Neg:Pos	26.6:1
Unhelp	Neg:Pos	26.3:1
Steal	Neg:Pos	25.3:1
Disappoint	Neg:Pos	24.8:1
Mice	Neg:Pos	24.4:1
Upset	Neg:Pos	24:1
Accuracy	0.7922	

Concluding remarks

This study aims to identify the determinants of backpackers' perception of security. More specifically, the impact of hostel's cleanliness, location, staff, atmosphere, and a guest's country of origin on backpacker's perception of security is studied. To respond to these research questions, 324,321 reviews were collected from Hostelworld.com and analyzed. Answers to these questions offer hostel managers a diagnostic tool to identify areas for further improvement in such a niche market that has experienced strong growth during the past decade. The analysis provides two main findings. First, hostel cleanliness, location, staff, and atmosphere have a positive and significant impact on backpackers' perceptions of security. These findings show that managers should be willing to invest in improving cleanliness, location, staff, and atmosphere to improve the security of their hostel. Additionally, using the CRA and Naïve Bayes Classifier, the main factors impacting security is identified. The result shows that how cleanliness (words like: Sheet, Room, Floor) and location (words like: Locate, Station, Area), staff (words like unhelpful and steal) and atmosphere (words like people and staff) impact backpackers' perceptions of security. Second, backpackers have a higher expectation for hostel security if they are coming from safer countries. This study provides evidence on how hostels' staff and atmosphere impact backpackers' perceptions of security.

Amblee (2015) showed that cleanliness followed by location determines backpackers' perception of security. In line with his result, this study also confirms that cleanliness followed by location is the main factors in determining bacpackers' perceptions of security. Additionally, our result further shows that staff and atmosphere also significantly impact bacpackers' perceptions of security. Compared to his result, our findings further show that backpackers coming from safer countries have a higher expectation for hostel security.

The practical implications of the present study are of various order. First, among all factors analyzed in this study (i.e., hostels' cleanliness, location, staff, atmosphere, and country of origin) cleanliness, staff and atmosphere are factors that hostel managers have control over.

Hostels can improve these aspects and make their guests feel more secure. Managers should be willing to invest in improving hostels' security by improving items such as locks, lockers, doors, and provide staff security training. What backpackers discuss online enable hotels' managers to take action, eliminate factors that cause low rating, in turn build strong brands. Other factors such as location and guest country of origin can be utilized in better planning, and targeting desired security levels. Suitable marketing mix strategy should be applied for different segments regarding the heterogeneity in terms of guest country of origin among international tourists. Second, online booking systems such as Hostelworld.com may extract valuable opinions from review content. Integrating automatic review mining with online search engines provides useful information about a certain hostel and facilitate hostel picking by potential guests.

Future research should explore how other backpackers' demographic characteristics such as gender, age, etc. impact backpackers' perceptions of security. Another direction could be studying backpackers' perceptions of security over time as backpackers' perception of security changes frequently. It would be interesting to compare and contrast findings between different time periods. Additional, it would also be interesting to study how hostels distance from a police station, hospitals and in general public facilities impact backpackers' perceptions of security.

References:

Amblee, N. (2015). The impact of cleanliness on customer perceptions of security in hostels: A WOM-based approach. *International Journal of Hospitality Management*, *49*, 37–39.

Brochado, A., & Rita, P. (2016). Exploring heterogeneity among backpackers in hostels. *Current Issues in Tourism*, 1–19.

Corman, S. R., Kuhn, T., McPhee, R. D., & Dooley, K. J. (2002). Studying complex discursive systems: Centering resonance analysis of communication. *Human Communication Research*, *28*(2), 157–206.

Cró, S., & Martins, A. M. (2017). The importance of security for hostel price premiums: European empirical evidence. *Tourism Management*, *60*, 159–165.

Cró, S., Martins, A. M., Simões, J. M., & Calisto, M. de L. (2018). Effect of Security on Hostels' Price Premiums: A Hedonic Pricing Approach. *Cornell Hospitality Quarterly*, 1938965518777224.

Dayour, F., Adongo, C. A., & Taale, F. (2016). Determinants of backpackers' expenditure. *Tourism Management Perspectives*, *17*, 36–43.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, *29*(2–3), 103–130.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, *1*(12).

Hecht, J.-A., & Martin, D. (2006). Backpacking and hostel-picking: an analysis from Canada. *International Journal of Contemporary Hospitality Management*, *18*(1), 69–77.

Hostelworld. (2016). *Annual Report*. Retrieved from http://www.hostelworldgroup.com/~/media/Files/H/Hostelworld/reports-and-presentations/hostelworld-annual-report-2016.pdf

Lima, R., Vicente, P., & others. (2017). Exploring the relationship between hostel service quality and guests' satisfaction: a comparison between younger and older guests. *E-Review of Tourism Research*, (5–6), 242–257.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415–463). Springer.

Maoz, D. (2007). Backpackers' motivations the role of culture and nationality. *Annals of Tourism Research*, *34*(1), 122–140.

Martins, M. R., Rachão, S., & Costa, R. A. da. (2018). Electronic word of mouth: Does it really matter to backpackers? Booking website reviews as an indicator for hostels' quality services. *Journal of Quality Assurance in Hospitality & Tourism*, 1–27.

Muzaini, H. (2006). Backpacking Southeast Asia: strategies of looking local. *Annals of Tourism Research*, *33*(1), 144–161.

Oliveira-Brochado, A., & Gameiro, C. (2013). Toward a better understanding of backpackers' motivations. *Tékhne*, *11*(2), 92–99.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 115–124).

Pang, B., Lee, L., & others. (2008). Opinion mining and sentiment analysis. *Foundations and Trends®in Information Retrieval*, *2*(1--2), 1–135.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143–157.

Richards, G., King, B., & others. (2003). Youth travel and backpacking. *Travel & Tourism Analyst*, (6), 1–23.

Shanahan, K. J., & Hyman, M. R. (2007). An exploratory study of desired hotel attributes for

American tourists vacationing in China and Ireland. *Journal of Vacation Marketing*, *13*(2), 107–118.

Shokoohyar, S. (2018). Ride-sharing platforms from drivers' perspective: Evidence from Uber and Lyft drivers. *International Journal of Data and Network Science*, *2*(4), 89–98.

Shortt, G., & Ruys, H. (1994). Hotel security: the needs of the mature age market. *International Journal of Contemporary Hospitality Management*, *6*(5), 14–19.

Teo, P., & Leong, S. (2006). A postcolonial analysis of backpacking. *Annals of Tourism Research*, *33*(1), 109–131.

Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, *2*(6), 282–292.

Wilson, A., Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2008). Services marketing: Integrating customer focus across the firm. *Berkshire: McGraw-Hill*.

Yang, E. C. L., Khoo-Lattimore, C., & Arcodia, C. (2017). A narrative review of Asian female travellers: Looking into the future through the past. *Current Issues in Tourism*, *20*(10), 1008–1027.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, *36*(3), 6527–6535.

Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, *24*(4), 720–734.

Zhang, J., Morrison, A. M., Tucker, H., & Wu, B. (2018). Am I a backpacker? Factors indicating the social identity of Chinese backpackers. *Journal of Travel Research*, *57*(4), 525–539.

Zhang, J., Tucker, H., Morrison, A. M., & Wu, B. (2017). Becoming a backpacker in China: A grounded theory approach to identity construction of backpackers. *Annals of Tourism Research*, *64*, 114–125.