Tourism Management 61 (2017) 43–54

Contents lists available at ScienceDirect

Tourism Management

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

Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis

M. Geetha [a, *], Pratap Singha [b], Sumedha Sinha [b]

[a] Department of Management Studies, IIT Madras, India
[b] IIT Madras, India

HIGHLIGHTS

- Sentiment analysis of online hotel reviews for explaining customer ratings.
- Premium and budget segment hotels in Goa considered for study.
- Statistically significant variation in ratings is explained by sentiment polarity.
- Sentiments are less positive for premium hotels than budget hotels in Goa.
- Premium hotels fair better in terms of staff performance and service.

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form
29 December 2016
Accepted 30 December 2016

Keywords:
Hotel categories
Customer ratings
Customer sentiments
Sentiment analysis

ABSTRACT

This study aims to establish a relationship between customer sentiments in online reviews and customer ratings for hotels. Customer sentiment refers to the emotions expressed by customers through the text reviews. These sentiments can be positive, negative or neutral. The study explores customer sentiments and expresses them in terms of customer sentiment polarity. Our results find consistency between customer ratings and actual customer feelings across hotels belonging to the two categories of premium and budget. Customer sentiment polarity explains significant variation in customer ratings across both the hotel categories. With regard to managerial implications, the study finds that, when compared with premium hotels, managers of budget hotels should improve their staff performance and hotel services. The present study is not exhaustive and other factors like customer review length and review title sentiment can be analyzed for their effects on customer ratings.

1. Introduction

Travel is an integral part of our lives. According to Pavaan Nanda, co-founder of Zostel, “Travel is seen as a mode of self-realization, exploration and experiencing different forms of lifestyles. Leisure travel is not a product of luxury but rather considered a necessity to consolidate one's energy.” (Tripzuki, 2013). In 2015 international tourism market grew by 4%, with 1184 million tourists travelling worldwide. This was led by 5% growth in Europe, the America, Asia and the Pacific. In 2016, international tourist arrivals are expected to grow by 4% worldwide (UNWTO, 2015). The Travel and Tourism Competitiveness Index Ranking (2015) indicate Europe (represented by 6 countries) in the top 10, making it the best continent for travel. Spain tops the ranking list helped by surge in international tourists coming from emerging countries like China, Brazil and Mexico. Emerging countries are also seeing an uptrend in the inflow of international tourists. Countries like Morocco and Saudi Arabia are upcoming attractions in the Middle East – North Africa region. Similarly, in the Sub-Saharan Africa region, Mauritius and Botswana are preferred by travelers. As far as cities are concerned, Hong Kong tops the list for travelers, closely followed by London (Euromonitor, 2016).

The dynamic landscape of global tourism industry is reflected through the advent of new destinations like mountain tourism in Korea (UNWTO, 2015). The tourism industry is also growing through non-conventional sub-divisions like indigenous tourism and eco-tourism. Countries like Australia and Canada have well developed indigenous tourism industries, popularly known as Aboriginal tourism (Aboriginal, 2016). Similarly, eco-tourism has

* Corresponding author.
E-mail address: m.geetha@iitm.ac.in (M. Geetha).

http://dx.doi.org/10.1016/j.tourman.2016.12.022
0261-5177/

been popular in countries like Costa Rica, Jordan and South Africa (Ecotourism, 2016).

For both host and the tourists' home countries, tourism industry helps to generate substantial economic benefits. Developing countries often promote themselves as tourism destinations to gain from the expected economic improvement (UNEP, 2016). Local economic development and enhanced tourism can only be realized with the hotel industry (Jones, Hillier, & Comfort, 2014). One such instance can be found in China, where the hotel industry is seeing increased revenue growth spurred by quick economic development, rising purchasing power, and reduced transportation costs (Xu, 2010; Zhang, 2011). Because of the micro nature of the hotels, they can generate employment through their labor intensive structure and can boost local spending quickly (Udemy Blog, 2014). So, it is imperative for a growing economy like India, to keep the hotels running efficiently and contributing to the country's Gross Domestic Product (GDP).

One way to improve the hotel industry is by better understanding of customers through ratings and reviews. Online customer ratings play a key role in the hospitality industry (Xie, Zhang, & Hang, 2014). Online hotel reviews also provide comparative and benchmarking insights about customer satisfaction (Mauri & Minazzi, 2013; Zhou, Ye, Pearce, & Wu, 2014). Online review websites that are dedicated to the rating of hotels have been gaining immense popularity (Buhalis & Law, 2008) because of increased impact of tourism, which contributes to 9.4% of global GDP (Baumgarten & Kent, 2010). Using consumer feedback, hotel ratings are assigned by few online websites. Ratings by these websites consider the number of reviews, age of the reviews, and quality (valence) of the reviews. Although, overall customer ratings are provided in these websites, there is a need to understand how far these ratings are consistent with the actual customer sentiments expressed through the reviews. Also this study needs to be performed across different domains of hotels to get better insights. Objective of our study is to analyze whether the perceived sentiments in opinions of the customers are consistent with the hotel ratings provided by them led to the following research questions:

Research question 1 Can customer review sentiments explain the customer ratings provided for hotels?

Research question 2 Is the relationship between customer review sentiments and hotel ratings are consistent across hotel categories of budget and premium?

2. Conceptual model and hypothesis development

Customer satisfaction and complaints are important for hotel performance (Assaf, Josiassen, Cvelbar, & Woo, 2015). Expectations of hotel guests when compared with actual experiences often fail to match due to various reasons. This lack of conformity can be enhanced by hotel star ratings and overall customer ratings (Rhee & Yang, 2015). Radojevic, Stanisic, and Stanic (2015) have found higher hotel ratings leading to increased customer satisfaction along with other attributes. Our study goes down further to a more granular level and tries to establish a relationship between the hotel ratings and the attributes desired by the customers through an understanding of customer sentiment present in online reviews.

One of the key elements to capture customer attention is through online review. Online review can be both positive and negative. Managing negative online reviews is imperative for managers for maintaining the image of the hotel. Study on the perceptions and evaluations of prospective customers toward an online negative review and any accompanying hotel response was done by Sparks, So, and Bradley (2015). Provision of an online response (versus no response) enhanced inferences that potential consumers draw regarding the business's trustworthiness and the extent to which it cares about its customers. Using a human voice and a timely response yielded favorable customer inferences. However, understanding how far negative a review is has not been defined in the study (Sparks et al., 2015). Our study provides a dimension to understand the sentiment polarity of a review in the context of being positive or negative, so that managers can better respond to such reviews.

Customer online rating patterns are domain specific. Travelers' rating patterns in the hotel review websites differ between independent and chain hotels across both profiles as well as regions (Banerjee & Chua, 2015). But, no study has ever tried to find out how far these ratings are consistent across the domains of different categories of hotels like premium and budget. In our present study, we have focused on implications of the hotel ratings and reviews across categories so that mangers can have separate set of strategies as per the hotel category.

Studies have been carried out to analyze whether customers of hotels that are certified for quality management are more satisfied than the customers of non-certified hotels of similar category and location. Findings show that quality certified hotels do not receive a statistically significantly better evaluation or rating from their customers and certified hotels have a statistically significantly lower rating in terms of value for money than non-certified hotels (Heras-Saizarbitoria, Arana, & Boiral, 2015). So, there is a need to understand what influences customers, beyond certifications, to give ratings. Customer satisfaction has significant effect in purchase intentions of customers in variety of sectors like banking, pest control, dry cleaning and fast food (Cronin & Taylor, 1992). In healthcare industry, patient satisfaction and service quality leads to higher probability of patient return to the same hospital in future (McAlexander, Kaldenburg, & Koenig, 1994). Customer satisfaction affects customer loyalty. Customer loyalty is an important factor for success of hotels. Customer loyalty is positively correlated with customer satisfaction and with hotel services like housekeeping, reception, food and beverage, and price (Kandampully & Suhartanto, 2000). A survey done by Bowen and Chen (2001) with 564 hotel customers from the hotel's database, found the relationship between customer loyalty and customer sentiment to be non-linear. The relationship between customer sentiment and loyalty rely on two pivotal points (Coyne, 1989). When customer satisfaction reaches the high point, loyalty increases dramatically thereafter. Similarly, beyond the lower point, customer loyalty decreases at the same rate (Oliva, Oliver, & MacMillan, 1992). So, understanding the right level of customer sentiment can help hotels leverage customer loyalty. Our study is a step in this direction. While customer satisfaction had significant effect on customer loyalty, customer's intentions of providing positive word-of-mouth about a hotel was less dependent on customer satisfaction (Getty & Thompson, 1995). Hotel service provider's attributes resulted in higher relationship quality with customers as well (Kim & Cha, 2002). Our study helps identify hotel services that might affect customer satisfaction levels both positively and negatively.

Business value of consumer reviews and management responses to hotel performance was studied by Xie et al. (2014). It has been found that overall rating, attribute ratings of purchase value, location and cleanliness, variation and volume of consumer reviews, and the number of management responses are significantly associated with hotel performance. In addition, variation and volume of consumer reviews moderate the relationship between overall rating and hotel performance. Also, a hotels' financial performance can be affected by customer online ratings. Some studies have focused on the impact of customer feedback on the revenue generation for hotels. In a study by Ye, Law, and Gu (2009) in which a mathematical model was developed to explain the impact of user-generated comments on hotel sales and profitability. Other researchers have explored the relationship of positive reviews and traffic to the business' website (Zhang, Ye, Law, & Li, 2011). A study by Torres, Singh, and Robertson-Ring (2015) explores the positive impact of a hotel's rating and number of reviews on the value generated through online transactions. Research demonstrate that online review website ratings as well as the number of reviews had positive relationship with the average size of each online booking transaction.

Studies had been done in tourism domain using Bing Liu's aspect-based opinion mining technique. Opinions available on the web as reviews were used to discover consumer preferences about tourism products, particularly hotels and restaurant by Marrese-Taylor, Velásquez, Bravo-Marquez, and Matsuo (2013). Results showed that tourism product reviews available on web sites contain valuable information about customer preferences that can be extracted using an aspect-based opinion mining approach. However, on average, the algorithms were only capable of extracting 35% of the explicit aspect expressions.

Sentiment analysis is already being adopted in different domains of research. Twitter sentiment analysis has been a topic of interest mainly because of its rapid growth as a popular platform for people to voice their opinion about a variety of topics. Tweet-level sentiment analysis is of two main types, supervised learning and lexicon based. SentiStrength software can be used along with a lexicon to identify the strength of sentiments in particularly informal texts (Saif, He, Fernandez, & Alani, 2015). Study to understand Tweet sentiments for hotels with sentiment analysis has been done by Philander and Zhong (2016). This study demonstrated how Twitter data could be analyzed through a cost-effective application to Las Vegas integrated-resort casino accounts. But a study on sentiment analysis on customer reviews for hotels and its subsequent relation with customer ratings is still lacking.

A study on the helpfulness and readership of online customer reviews on Amazon website has been done by Salehan and Kim (2015). A sentiment mining approach was used to find out how the underlying sentiments of customer reviews, along with other variables, determine the importance of reviews. But, no study has focused on finding how such reviews can affect the online ratings given by customers. Our study aims to bridge this gap by relating online ratings with customer review sentiment. Online ratings given to hotels can be considered consistent when they are related to the underlying customer sentiments of the reviews. Hence, the key variable for our study is customer sentiment polarity. Customer Sentiment Polarity is defined as the ratio between the number of positive words and the number of negative words in a given review. Customer sentiment refers to the total amount of sentiment that exists in a text, both positive and negative. And polarity refers to the direction of this sentiment in terms of being positive, negative or neutral (Salehan & Kim, 2015). Polarity detection can be done at the word level (Sayeed, Graber, Rusk, & Weinberg, 2012). In the word level, each word in the text is analyzed and classified (Lima, Castro, & Corchado, 2015). The tagging of a word as positive or negative was done by comparing them with a lexicon of words present in the “Sentiment” package in R using the Naïve Bayes algorithm. We measured the effect of this independent variable on the dependent variable, i.e., customer ratings. Online customer ratings represent the customers' satisfaction with the product (Engler, Winter, & Schulz, 2015). Study by Engler et al. (2015) establishes pre-purchase expectations and actual product performance as determinants of ratings. Online ratings can be helpful as they reduce the uncertainty about the products (Chen & Xie, 2008). But, how online ratings can be used for hotels still not addressed adequately. Our study explores how these customer ratings are affected by review sentiment polarity.

Accordingly we propose the following hypothesis:

H_1: Customer sentiment polarity has a positive effect on customer rating.

The research model can be summarized as follows in Fig. 1:

3. Methodology

3.1. Sentiment analysis

Data in the form of texts is one of the major forms of unstructured data. It forms a large repository of information provided we can extract the right content from it. One way to do this is through sentiment analysis. Sentiment Analysis has emerged as an important aspect in text analytics. Sentiment is an attitude, thought, or judgment prompted by feeling and sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities (Fang & Zhan, 2015). When compared with more traditional market research methods (e.g. surveys or opinion polls), sentiment analysis has the advantage of being more cost and time efficient. Also, it is a nonintrusive method to extract consumers' opinions and sentiments in real-time—avoiding recall biases generally associated with post consumption self-report measurements (Rylander, Propst, & McMurtry, 1995).

Sentiment analysis has been performed on various domains like twitter and amazon reviews to gauge customer sentiments. But, a comprehensive study of customer sentiments in the hotels domain is still lacking. Most analyses have focused on analysis through ready-made online software. No inherently generated algorithm has been used so far. Our study aims to address these issues.

3.2. Location for data collection

Tourism industry in India has evolved into a thriving domain. In tourism, there was a rise of nine per cent during last year and 14 per cent in the previous year. India ranks 7th in numbers of World Heritage sites (World Economic Forum, 2007). Goa is one of the major tourist spots in India (Touropia, 2015; About Travel, 2016). Goa is poised to have 10 million tourists arriving by 2017.

The Investment Promotion Board (IPB) in India has approved the proposal for 11 projects including five-star hotels, with an investment of about Rs. 1000-crore in Goa (The Economic Times, 2015). Goa Tourism would soon launch an e-commerce site to enable domestic and foreign tourist book online their boarding and lodging, as well as, activities like adventure and water sports, music festival tickets, restaurants and river cruises (NDTV, 2015).

With all these initiatives in place, Goa is becoming one of the most important destinations in the Indian tourism landscape. To stay relevant in today's connected world, hotels in Goa need to maneuver their online presence and resources to attract tourists and provide better service.

The distribution of hotels in GOA is as follows:

There are a total of 496 hotels in Goa, out of which 110 belong to the budget category. Premium category hotels include both luxury and resort type hotels and are 138 in number. The remaining 248 hotels belong to other categories including mid-range (Tripadvisor, 2016).

Some of the well-known hotel brands finding their presence in Goa are Clarion, Golden Tulip, Hyatt, Marriott, Vivanta and Taj Exotica. Apart from hotels, other popular staying options available for travelers are guest houses, specialty lodging and holiday rentals. These are well distributed throughout popular destinations in Goa like Panaji, Calangute, Candolim, Vasco Da Gama, Dona Paula and Baga. For the non-conventional tourists, Goa offers unique heritage

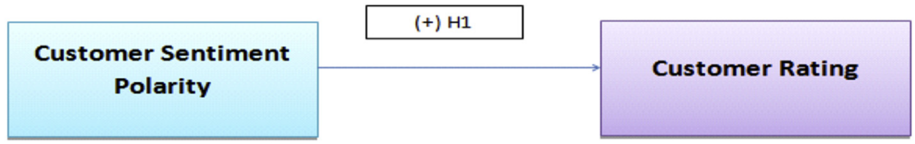

Fig. 1. Conceptual model.

and boutique stays, backpacker & adventure stays and beach huts, tents and cottages (Planet Goa, 2016).

3.3. Sampling

Simple Random Sampling (SRS) was done for selecting a sample of 20 hotels from each of budget category and premium category. For both our focus categories, the sampling frame was known. For budget category it is 110 hotels and for premium category it is 138 hotels. These frames comprise the finite populations for each category respectively. Simple random sampling for each category was done as follows:

Firstly, a MS Excel 2010 spreadsheet was selected. For a given category, all the hotels belonging to the sampling frame were placed in the first column of the spreadsheet. The adjacent column was populated with random numbers generated with rand() function in MS Excel 2010. Then, the rows of the spreadsheet were sorted using the random numbers in the second column. Finally, the hotels in the first 20 rows of the resulting spreadsheet were selected as the simple random sample for the given category (Stine & Foster, 2014).

For budget category in Goa, the total hotel population is 110, out of which we have drawn a sample of 20. Similarly, for premium category in Goa, the total hotel population is 138, out of which our sample has 20 hotels. A SRS requires that sample size should be less than 10% of the population size. For our analysis, this condition is not met for both the categories. To compensate for this, we multiple the standard errors with finite population correction (FPC) factors, in the subsequent regression analysis, thus narrowing the respective confidence intervals. FPC is defined as follows:

FPC = [(N-n)/(N-1)]^0.5 where, N is the population size and n is the sample size (Stine & Foster, 2014;; Zikmund, Babin, Carr, & Griffin, 2009, p. 435). FPC for budget and premium categories are found to be 0.91 and 0.93 respectively.

The sample size condition for extending the regression analysis inferences from samples to the entire populations requires that n > 10 * (Absolute Value of Kurtosis of the model residuals), where, n is the sample size, i.e., equal to the number of residuals (Stine & Foster, 2014).

For Budget Category, $n = 20$ and $|\text{Kurtosis}| = 0.853$. So, $n > 10 * |\text{Kurtosis}|$. Similarly, for Premium Category, $n = 20$ and $|\text{Kurtosis}| = 0.144$. So, again $n > 10 * |\text{Kurtosis}|$. This meets the sample size condition for both budget and premium category. Also, with sample size 20, the margins of error for regression slopes for budget and premium categories are coming out to be 0.02 and 0.01, which are within acceptable ranges (Stine & Foster, 2014;; Zikmund et al., 2009, p. 435). We can thus generalize the results of regression analysis for the focus hotel categories listed online in Goa.

3.4. Data analysis

Different classification techniques have been previously used for classifying words as per underlying sentiment.

Pang, Lee, and Vaithyanathan (2002), Ye, Zhang, and Law (2008) and Kang, Yoo, and Han (2009) applied a traditional topic-based document classification method to classify a review document as positive or negative. Dave, Lawrence, and Pennock (2003) had used a score function formula to determine positive and negative sentiments stochastically. Myung, Lee, and Lee (2008) used a syntax analyzer for product review analysis and extracted a sentiment word candidate from a predicate.

Use of a lexicon has been found perennial in any classification technique. Sista and Srinivasan (2004) in their study focused on classification experiment of movie reviews through the use of a lexicon constructed using the General Inquirer Lexicon and the WordNet database. The lexicon was used as part of the Naïve Bayes algorithm. Fahrni and Klenner (2008) constructed a self-made lexicon with a two-stage model, and the classification performance was good when the self-constructed lexicon and SentiWordNet were used simultaneously. Cho and Lee (2006) manually constructed a lexicon in order to find eight kinds of basic emotions in a person from the lyrics of songs, and performed a classification experiment using a supervised learning algorithm. For our study, we have considered a lexicon of positive and negative words, as used by Hu and Liu (2004) and Liu, Hu, and Cheng (2005).

For our classification algorithm, we have used the Naïve Bayes classification technique. We have chosen this algorithm among various other algorithms because of its demonstrated success. Miao, Li, and Dai (2009) measured the ranking of reviews by analyzing the quality of an opinion with a study on the sentiment mining and retrieval system using the Naïve Bayes algorithm. Tan and Zhang (2008) classified Chinese documents as positive and negative. Here, Naïve Bayes was used as one of the classification techniques. Naïve Bayes uses the lexicon of words to match words from documents against the lexicon. Then it assigns each word the probability of being positive or negative, independent of other words.

But, before the above mentioned techniques can be employed, the text data needs to be processed and made ready for analysis. The various steps involved are as follows (see Figs. 2 and 3):

All the reviews per hotel were collected column-wise in MS Excel and the files were saved in CSV (Comma Delimited) format for easy readability in the R environment. This was followed by the data cleaning step using "tm" package and "SnowballC" package in R.

After the data was cleaned, all the words were loaded in a term document matrix. Matrix sparsity was reduced to 10% to remove all irrelevant terms.

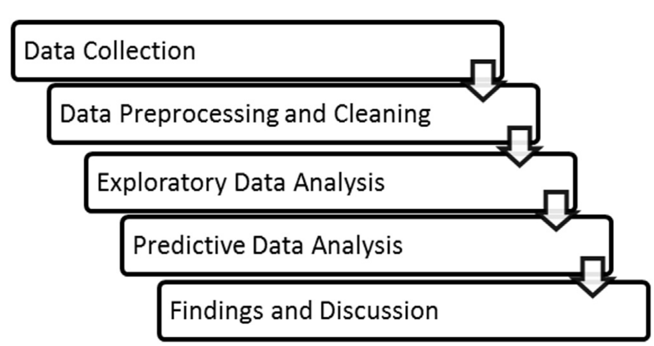

Fig. 2. Steps of analysis.

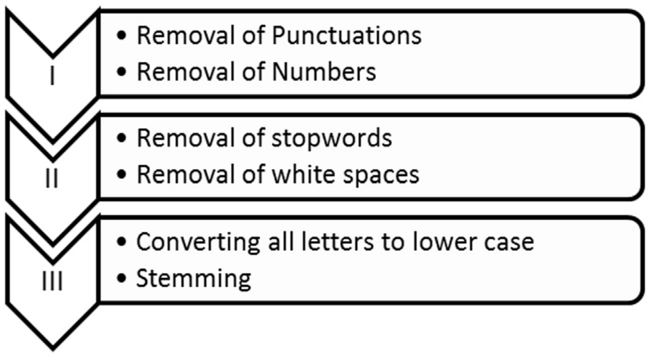

Fig. 3. Data cleaning process.

4. Findings and discussion

4.1. Exploratory data analysis

As part of the exploratory data analysis, we looked at most common words in reviews in both the categories. The top terms included words like “hotel”, “good”, “staff”, “service”, etc. The results are shown in Figs. 4–7:

Wordclouds provide better visual representation and help to make comparisons between the two categories. Higher the usage frequency of a word, larger will its presence in the word cloud. From the above charts and wordclouds, it is evident that certain terms have been used more frequently in case of premium category hotels than for budget category hotels. But, just the mere difference in frequency in the usage of these terms does not allow us to draw sufficient conclusion regarding each category of hotels. We must look at the correlations among the terms and how frequently they have been used with other terms so that we can draw some meaningful results. We did this through the usage of cluster analysis. Clustering techniques have been used previously to detect polarity in a supervised way (Cambria, Mazzocco, Hussain, & Eckl, 2011). Here we have used clustering as an unsupervised statistical learning method. In unsupervised learning, we do not have any associated response variable. So, the aim is not prediction, but to find interesting patterns and insights within the data. Some of the questions we wished to answer through this analysis were whether there were informative ways to visualize the data and discover subgroups among the words (James, Witten, Hastie, & Tibshirani, 2016). Using the Euclidean distance matrix, being represented in

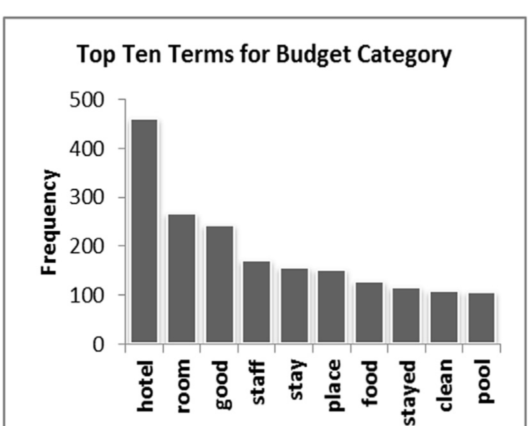

Fig. 4. Top ten terms for budget category.

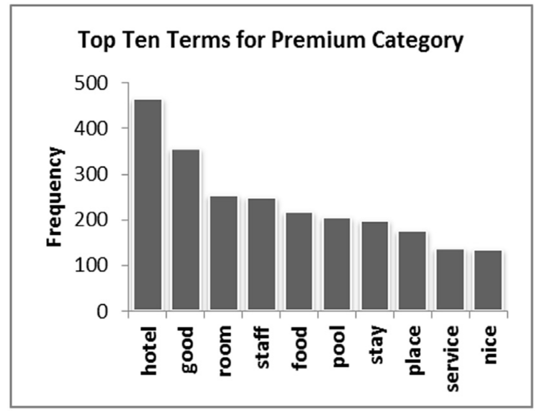

Fig. 5. Top ten terms for premium category.

Fig. 6. Wordcloud for budget category.

Fig. 7. Wordcloud for premium category.

terms of correlation between terms, hierarchical clustering was performed across both the categories. This analysis allowed us to see the different clusters formed by terms based on how strongly they are correlated in terms of usage, and how the smaller clusters in turn combined to form bigger clusters see (Figs. 8 and 9).

One of the key requirements for understanding our research questions is the understanding the customer sentiments hidden in the reviews. The cluster analysis helps us in understanding these sentiments better by looking at the terms used by the customers in their reviews. The Height of the cluster dendrograms is inversely proportional to the correlation between terms. Highly correlated terms form cluster at lower points of the dendrogram. Also, two words at the same height, but belonging to different clusters, will have very far correlations.

4.2. Predictive data analysis

To test our hypothesis, we used simple linear regression.

The regression model tries to determine whether a linear relationship exists between customer rating and customer sentiment polarity. Here, customer sentiment polarity is the independent variable and customer rating is the dependent variable. We wish to find out whether a linear change in customer sentiment polarity leads to a linear change in customer rating. Customer Rating is obtained directly from the website.

Customer Sentiment Polarity = Positive Words/Negative Words of the reviews

Polarity is "positive" when the above ratio is greater than 1.5, "neutral" when it is between 1 and 1.5 and "negative" when it is 0 to less than 1. To determine the polarity and sentiment of each review, Naïve Bayes Classification algorithm was used against a lexicon of words. The algorithm was already trained against an extensive repository of sentiment based words. It comes as default in the "sentiment" package in R. Figs. 10–15 provide insights into emotions and sentiment polarities for the hotels considered in sample.

As we can see from the above plots, for budget category hotels, negative emotions and negative sentiments are more than those for premium category hotels.

Result of regression analysis is summarized in Figs. 16 and 17:

From the above regression analysis for budget category hotels, we find that the value of the t statistic for the slope is 3.77. This is greater than the critical t value of 2.1 for $\alpha = 5\%$ and degrees of freedom 18. Thus we find support for the linear relationship between customer sentiment polarity and customer rating. The standard errors for intercept and slope are multiplied with FPC factor of 0.91 to compensate for sample size being greater than 10% of the population. So, the actual standard errors are 0.19 and 0.009 respectively. The coefficient of determination or R square has a value of 0.44. So, 44% of the variation in the customer ratings is explained by customer sentiment polarity.

Similarly, from the regression analysis for premium category hotels, we find that the value of the t statistic for the slope is 2.21. This is greater than the critical t value of 2.1 for $\alpha = 5\%$ and degrees of freedom 18. Thus we find support for the linear relationship between customer sentiment polarity and customer rating. The standard errors for intercept and slope are multiplied with FPC factor of 0.93 to compensate for sample size being greater than 10% of the population. So, the actual standard errors are 0.15 and 0.0055 respectively. The coefficient of determination or R square has a value of 0.21. So, 21% of the variation in the customer ratings is explained by customer sentiment polarity.

The linear relationship between customer sentiment polarity and customer rating is well established by the regression model. And, this is found true across both the categories of budget and

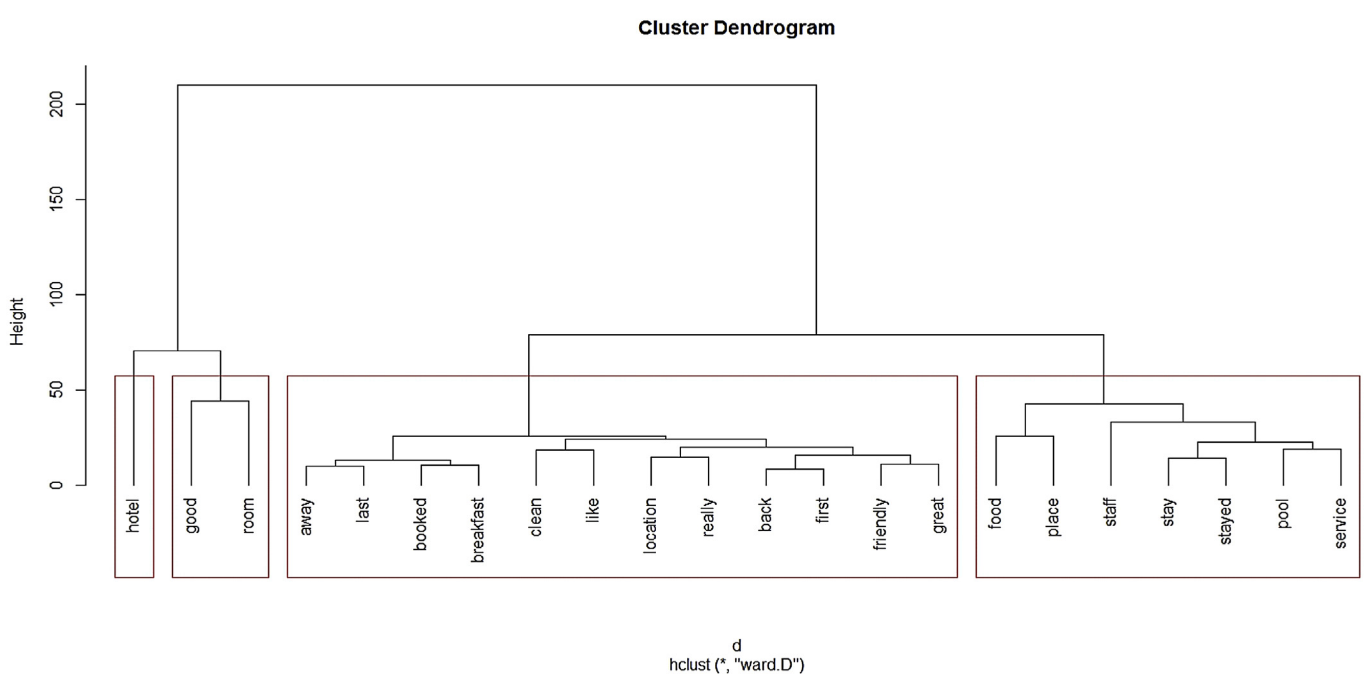

Fig. 8. Cluster dendrogram for budget category.

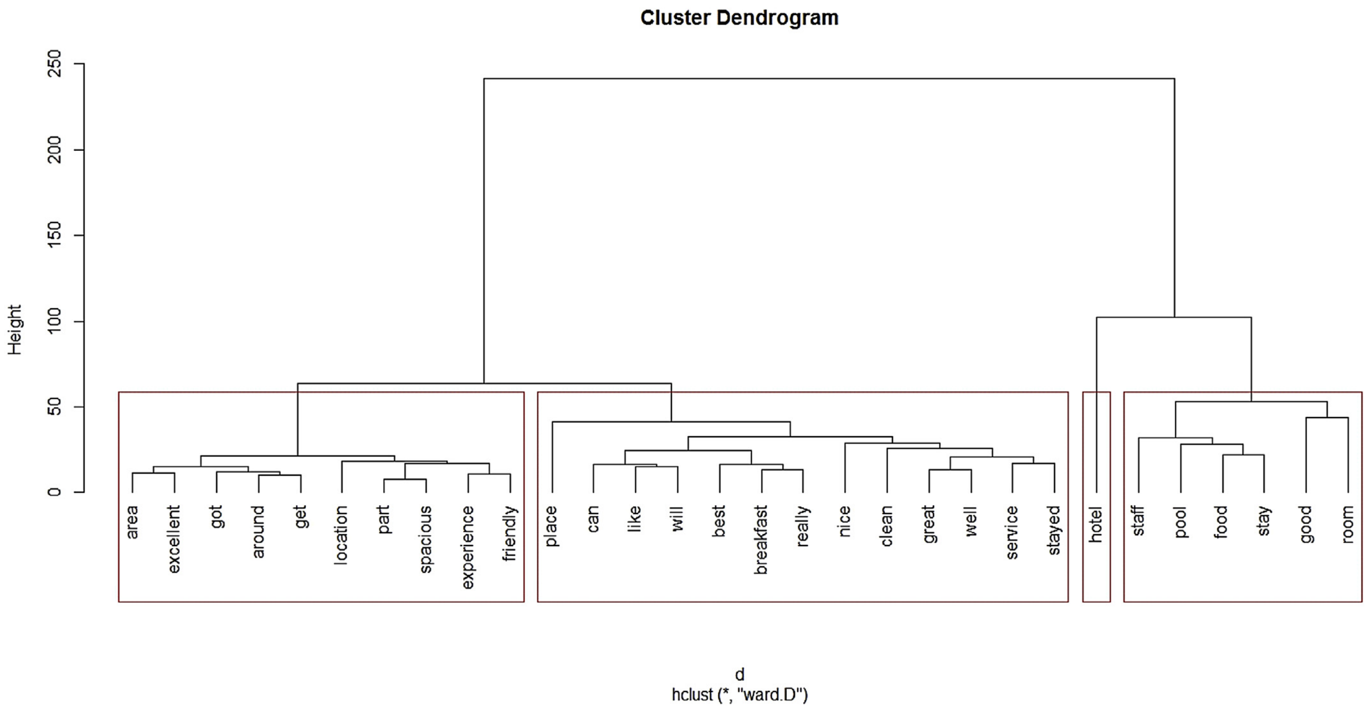

Fig. 9. Cluster dendrogram for premium category.

Sr. No.	*Hotel*	*Category*	*Emotions (Out of 20)*							*Polarity (Out of 20)*				
			Joy	Sadness	Anger	Fear	Surprise	Disgust	Unknown	Positive	Negative	Neutral	Mean Polarity	Ratings
1	**A**	**Budget**	**18**	**1**	**0**	**0**	**0**	**0**	**1**	**15**	**0**	**5**	**37.32**	**4**
2	**B**	**Budget**	**19**	**1**	**0**	**0**	**0**	**0**	**0**	**12**	**1**	**7**	**33.67**	**4.5**
3	**C**	**Budget**	**17**	**0**	**2**	**0**	**0**	**0**	**1**	**17**	**1**	**2**	**54.42**	**5**
4	**D**	**Budget**	**16**	**1**	**0**	**0**	**0**	**1**	**2**	**16**	**2**	**2**	**34.45**	**3.5**
5	**E**	**Budget**	**20**	**0**	**0**	**0**	**0**	**0**	**0**	**16**	**0**	**4**	**41.24**	**4**
6	**F**	**Budget**	**17**	**1**	**2**	**0**	**0**	**0**	**0**	**15**	**2**	**3**	**16.58**	**4**
7	**G**	**Budget**	**18**	**1**	**0**	**0**	**0**	**0**	**1**	**13**	**4**	**3**	**20.71**	**4**
8	**H**	**Budget**	**19**	**0**	**0**	**0**	**0**	**1**	**0**	**13**	**2**	**5**	**16.12**	**4.5**
9	**I**	**Budget**	**13**	**0**	**4**	**1**	**0**	**0**	**2**	**15**	**3**	**2**	**16.3**	**4**
10	**J**	**Budget**	**15**	**0**	**2**	**0**	**1**	**0**	**2**	**16**	**2**	**2**	**32.9**	**3.5**
11	**K**	**Budget**	**9**	**2**	**1**	**0**	**0**	**0**	**8**	**10**	**7**	**3**	**16.48**	**4.5**
12	**L**	**Budget**	**9**	**4**	**1**	**1**	**1**	**1**	**3**	**6**	**5**	**9**	**3.48**	**3**
13	**M**	**Budget**	**13**	**5**	**2**	**0**	**0**	**0**	**0**	**2**	**13**	**5**	**4.72**	**2.5**
14	**N**	**Budget**	**11**	**2**	**4**	**2**	**0**	**0**	**1**	**6**	**12**	**2**	**6.72**	**3**
15	**O**	**Budget**	**4**	**3**	**0**	**1**	**1**	**1**	**10**	**9**	**7**	**4**	**13.9**	**3**
16	**P**	**Budget**	**9**	**5**	**0**	**0**	**0**	**0**	**6**	**5**	**11**	**4**	**12.07**	**4**
17	**Q**	**Budget**	**16**	**1**	**1**	**1**	**1**	**0**	**0**	**8**	**5**	**7**	**13.65**	**3**
18	**R**	**Budget**	**13**	**1**	**3**	**1**	**0**	**0**	**2**	**5**	**8**	**7**	**4.99**	**3**
19	**S**	**Budget**	**11**	**1**	**1**	**0**	**0**	**0**	**7**	**10**	**8**	**2**	**10.39**	**3**
20	**T**	**Budget**	**16**	**0**	**1**	**2**	**0**	**1**	**0**	**10**	**6**	**4**	**11.8**	**3**

Fig. 10. Emotions and sentiments for budget category hotels.

premium. Our first research question was to understand whether customer review sentiment explains the customer ratings provided for hotels in Goa. It is evident from the study that customer sentiment explains 44% of the variation in customer rating for budget category and 21% of the variation in customer rating for premium category. So, the results support that:

Customer review sentiments explain the customer ratings provided for hotels.

As far as research question 2 is concerned, we found the following inferences:

Sr.No.	Hotel	Category	Emotion (Out of 20)							Polarity (Out of 20)				
			Joy	Sadness	Anger	Fear	Surprise	Disgust	Unknown	Positive	Negative	Neutral	Mean	Ratings
1	A	Premium	15	0	0	0	4	0	1	15	0	5	19.03	4.5
2	B	Premium	17	0	1	0	1	0	1	10	5	5	9.37	3.5
3	C	Premium	20	0	0	0	0	0	0	16	0	4	9.22	4
4	D	Premium	19	1	0	0	0	0	0	14	3	3	27.72	4
5	E	Premium	19	0	0	0	1	0	0	19	0	1	31.35	5
6	F	Premium	19	0	0	0	1	0	0	14	1	5	12.45	4
7	G	Premium	17	0	1	0	1	0	1	17	0	3	30.19	4.5
8	H	Premium	18	0	1	0	1	0	0	17	1	2	44.82	4.5
9	I	Premium	18	0	0	0	1	0	1	16	0	4	15.22	4
10	J	Premium	14	2	1	1	1	0	1	11	3	6	14.93	4
11	K	Premium	15	0	1	0	0	0	4	19	1	0	45.9	4
12	L	Premium	18	0	1	0	0	0	1	13	0	7	25.54	3.5
13	M	Premium	16	0	0	0	2	1	1	15	1	4	13.15	4
14	N	Premium	15	0	0	1	3	0	1	15	1	4	51.6	4.5
15	O	Premium	17	0	1	0	0	0	2	12	2	6	16.51	4.5
16	P	Premium	19	0	1	0	0	0	0	14	2	4	19.95	4.5
17	Q	Premium	15	3	0	0	0	0	2	15	1	4	11.37	4
18	R	Premium	19	0	0	0	1	0	0	14	1	5	11.18	3.5
19	S	Premium	15	0	0	1	1	0	3	12	1	7	10.13	4
20	T	Premium	20	0	0	0	0	0	0	11	3	6	16.47	4

Fig. 11. Emotions and sentiments for premium category hotels.

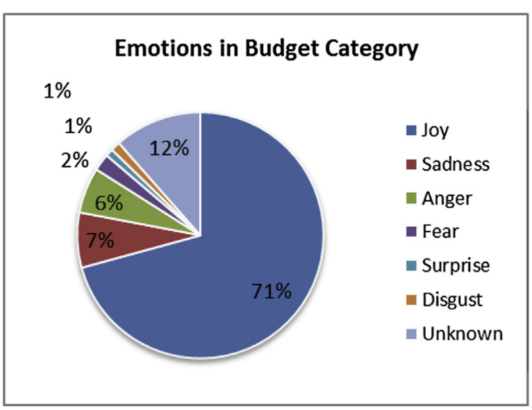

Fig. 12. Emotions in budget category.

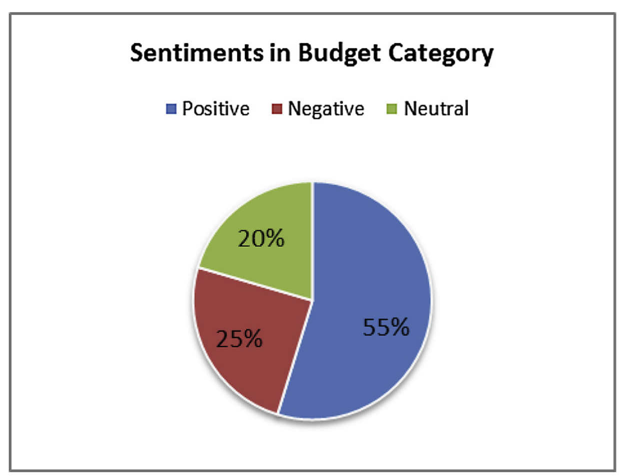

Fig. 13. Sentiments in budget category.

When we compare both the categories of hotels, we find that the validity of the regression model is much stronger for budget category with a p-value of 0.0014, while for premium category it is only 0.04. Again, the role of customer sentiment polarity in explaining customer rating is much more for budget category hotels, whereby the variation is explained by 44%. For premium category, this is only by 21%. But, at the same time, the deviation from the mean predicted ratings is much stronger in budget category than in premium category. This is observed from the standard error values for each category. For budget category this is 52%, while for premium category it is 36%. If we consider variation across the model fit as one of the measures of goodness of fit, then premium category seems to be doing better. So, relationship between customer review sentiments and hotel ratings are not entirely consistent across hotel categories of budget and premium.

Our present study finds support from similar studies done on a variety of domains. Turney (2002) could classify reviews from automobiles, banks, movies and travel destinations into recommended and non-recommended categories using unsupervised machine learning algorithm. Extending further, study done by Pang and Lee (2005) focused on the relationship between sentiments hidden in movie reviews and rating granularity on a five point scale. Using a meta-algorithm, sentiments could accurately be represented through a rating scale and the performance was better than different versions of SVM (Support Vector Machines). Along with sentiment orientation of movie reviews, Thet, Na, and Khoo

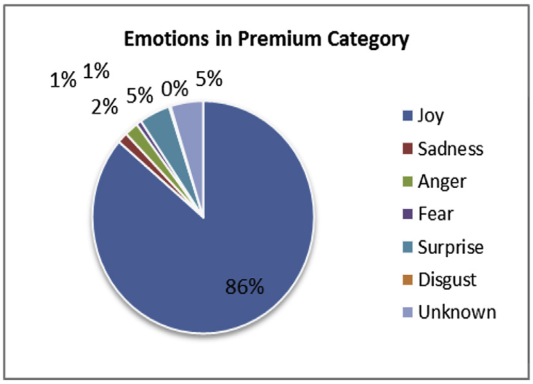

Fig. 14. Emotions in premium category.

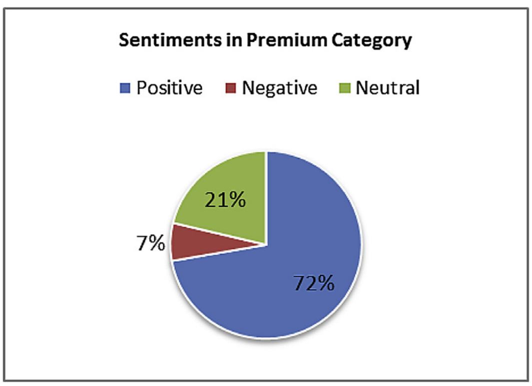

Fig. 15. Sentiments in premium category.

Regression Statistics for Budget Hotels					
R Square	**0.44**				
Standard Error	**0.52**				
ANOVA					
	Df	***SS***	***MS***	***F***	***Significance F***
Regression	**1**	**3.99**	**3.99**	**14.22**	**0.00**
Residual	**18**	**5.06**	**0.28**		
Total	**19**	**9.05**			
	Coefficients	***Standard Error***	***t Stat***	***P-value***	
Intercept	**2.99**	**0.21**	**14.24**	**0.00**	
Mean	**0.03**	**0.01**	**3.77**	**0.00**	

Fig. 16. Regression output for budget category.

(2010) could bring out the sentiment strength of the reviewer towards various aspects of the movie. Parallels with our study could be drawn whereby we have established similar relationships combining both unsupervised (cluster analysis) and supervised (least square regression) learning techniques for hotel reviews. In the e-commerce domain, Hu and Liu (2004) were able to summarize online customer reviews of products using text mining and sentiment classification. As done in our study, reviews were classified based on sentiment polarities of positive and negative, bringing more insights both for customers and hoteliers. On similar lines, sentiment analysis has been done on web searches of items by Dave et al. (2003), providing clarity on customer sentiments associated with each item attribute. In the microblogging domain, sentiment analysis has been done using Twitter as the corpus (Pak & Paroubek, 2010). Tumasjan, Sprenger, Sandner, and Welpe (2010) have performed sentiment analysis on Twitter blogs for predicting political outcomes. On social media platforms like Facebook, sentiment analysis of messages provide opportunity for personalized e-learning and subsequent feedback for teachers and rating of courses (Ortigosa, Martín, & Carro, 2014). Troussas, Virvou, Espinosa, Llaguno, and Caro (2013) performed sentiment analysis of Facebook status messages using the Naïve Bayes Classifier, similar to the classification technique used in our present study.

5. Managerial implications

In a study by Xie et al. (2014), customer reviews and ratings are of great business value to hotels. Hotel sales and profitability were linked to customer reviews (Ye et al., 2009). A study by Torres et al. (2015) demonstrated that hotel ratings and reviews on online sites influence positively the size of online transactions done with regard to hotel bookings.

Our study can help managers look beyond the hotel ratings into the sentiments customers are having regarding the hotels. Human language can express emotions which quantitative ratings cannot capture. With the help of the sentiment analysis done on the customer reviews in our present study, managers can look into these granular details and better respond to customer needs.

From the exploratory analysis itself, it is clear that customers are having better sentiments with regard to premium hotels than budget hotels. In the most frequent terms used for both the categories, we see that the term “good” has been used 355 number of times for premium category hotels while for budget category hotels, it has been used 243 number of times. From the wordclouds, a comparative study of the words used for the categories reveal certain directions for managers to look at. The word “staff” has been referred to a more frequently in premium category hotels than in budget category hotels. The term “service” finds a mention in the premium category but not in the budget category.

Looking at the hierarchical cluster dendrogram plots, we find the term “staff” has been used more frequently with the term “good” for premium category hotels than for budget ones. This is evident from the closer clusters they form. This implies that though the premium hotels are doing a good enough job with regard to

Regression Statistics for Premium Hotels	
R Square	**0.21**
Standard Error	**0.36**

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	**1**	**0.63**	**0.63**	**4.90**	**0.04**
Residual	**18**	**2.31**	**0.13**		
Total	**19**	**2.94**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**3.82**	**0.16**	**23.93**	**0.00**
Mean	**0.01**	**0.01**	**2.21**	**0.04**

Fig. 17. Regression output for premium category.

ensuring a good service from their staff, the budget category hotel managers need to look at ways to improve their staffing service. From further analysis of the hierarchical clusters, we observe that the term "service" appears in close clusters with the terms "great" and "well" for premium category hotels. But, for budget category hotels, the term shows no such correlation. So, managers in budget category hotels in Goa need to concentrate more on improving their service.

In the emotional classification of the reviews, we find that for the budget category hotels, the emotions are more negative and granular. Also, the ratio of number of positive to negative reviews is more for premium category hotels than in budget category hotels. Managers should be wary of such sentiments as word of mouth communications have lasting effects on hotel images. Westbrook (1987) defined word-of-mouth as: "informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services and/or their sellers". Word of mouth communication has the potential to influence consumer purchase decisions, customer acquisition, and consequently results in increased revenue for organizations (Litvin, Goldsmith, & Pan, 2008; Trusov, Bucklin, & Pauwels, 2009). Responding to any such negative reviews is imperative for the hotel management (Sparks et al., 2015). Our study can help managers in dealing with negative reviews promptly by understanding them in a better way.

From the regression analysis, we find that the variation in hotel ratings is explained by the customer sentiment to a greater extent in case of budget category hotels than in case of premium hotels. So, it is possible that factors other than customer ratings are affecting the customer ratings for the premium hotels. Surveys can be conducted by the hotel management along with sentiment analysis to better understand customer grievances. Often, fake reviews on websites can tarnish the image of the hotels. Managers should find tools to manage these reviews.

Customer satisfaction is correlated with hotel services like housekeeping, reception, food and beverage, and price (Kandampully & Suhartanto, 2000). As hotel managers can review their services to improve customer loyalty, customers can also choose from services that brings them desired satisfaction. Our study enables customers to look beyond online customer ratings while selecting hotels in Goa. It brings forth experiences of previous customers in a qualitative and interpretable manner, so that new customers can make informed decisions. Going through several thousands of comments present online for hotel reviewing purpose could be a tedious task. Our present study summarizes the underlying sentiments of the comments for customers to easily comprehend and decide. In addition, through our cluster analysis, customers can make deep dives into specific attributes of services provided by hotels in the two categories of budget and premium. Sometimes, fake ratings can distort the actual image of the hotels for the customers. Our study presents a framework for relating ratings with reviews, which can be used for validation. Differences exist between what hotel mangers perceive as customer value and what customer's actual experiences highlight. There is a need to align these differences in perception to optimize value delivered (Nasution & Mavondo, 2008). Our study provides a common ground for such alignment.

6. Future research and limitations

While our research has valuable contributions, it also has some limitations. The present study successfully concludes that there exists a positive effect of customer sentiment polarity on customer ratings for hotels across categories. But, at the same time, this influence is not totally responsible for customer ratings. Other factors like title sentiment and review length could be possible explanations as well. Future research could address this.

Secondly, we have used a lexicon of positive and negative words to run the Naïve Bayes algorithm for sentiment analysis. The lexicon is not an exhaustive list and the algorithm can be trained further with better contexts focusing on the hotel industry.

Thirdly, although the Naïve Bayes algorithm is tried and tested, it has its limitations. As it treats each word independently of the other, it may miss out on a hidden sentiment that is being expressed by the word. So, sometimes, looking at words alone is not enough. The underlying topic needs to be looked at as well. Other algorithms like POS Tagging can be used for finding out customer sentiment. Also, other platforms like NTLK in Python can be used for analysis.

Finally, we have considered reviews from only one website. This might be prone to bias. So, future studies can be conducted by collecting reviews from different websites for the same set of hotels.

References

Aboriginal (2016). Available from: http://www.tourism.australia.com/aboriginal.aspx, https://aboriginalcanada.ca/corporate/, Accessed 09.07.16.
About Travel (Jan, 2016). Available from: http://goindia.about.com/od/topdestinations/tp/10-Must-See-India-Destinations.htm, Accessed 19.02.16.
Assaf, A. G., Josiassen, A., Cvelbar, L. K., & Woo, L. (2015). The effects of customer

voice on hotel performance. *International Journal of Hospitality Management, 44*(2015), 77–83.
Banerjee, S., & Chua, A. Y. K. (2015). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management, 53*(2016), 125e131.
Baumgarten, J. C., & Kent, G. J. W. (2010). *Executive summary: Travel and tourism economic impact*. London, United Kingdom: World Travel and Tourism Council.
Bowen, J. T., & Chen, S. L. (2001). The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management, 13*(5), 213–217.
Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the internetethe state of eTourism research. *Tourism Management, 29*(4), 609e623.
Cambria, E., Mazzocco, T., Hussain, A., & Eckl, C. (2011). Senticmedoids: Organizing affective common sense knowledge in a multi-dimensional vector space. *LNCS, 6677*, 601–610.
Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science, 54*(3), 477–491.
Cho, Y., & Lee, K. (2006). Automatic affect recognition using natural language processing techniques and manually built affect lexicon. *IEICE Transactions on Information & Systems*, 2964–2971.
Coyne, K. (1989). Beyond service fads-meaningful strategies for the real wor. *MIT Sloan Management Review, 30*(4), 69.
Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *The Journal of Marketing*, 55–68.
Dave, K., Lawrence, S., & Pennock, D. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international world wide web conference (WWW '03)*.
The Economic Times (Dec, 2015). Available from: http://articles.economictimes.indiatimes.com/2015-12-15/news/69061782_1_goa-government-chief-minister-laxmikant-parsekar-20-crore; Accessed 16.02.16.
Ecotourism (2016). Available from https://en.wikipedia.org/wiki/Ecotourism; Accessed 09.07.16.
Engler, T. H., Winter, P., & Schulz, M. (2015). Understanding online product ratings: A customer satisfaction model. *Journal of Retailing and Consumer Services, 27*(2015), 113–120.
Euromonitor. (2016). *Top 100 city destination ranking*. Available from: http://blog.euromonitor.com/2016/01/top-100-city-destinations-ranking-2016.html Accessed 08.07.16.
Fahrni, A., & Klenner, M.. (2008). Old wine and warm beer: Target-specific sentiment analysis of adjectives. In Proceedings of the AISB.
Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data, 2*(2015), 5. http://dx.doi.org/10.1186/s40537-015-0015-2.
Getty, J. M., & Thompson, K. N. (1995). The relationship between quality, satisfaction, and recommending behavior in lodging decisions. *Journal of Hospitality & Leisure Marketing, 2*(3), 3–22.
Heras-Saizarbitoria, I., Arana, G., & Boiral, O. (2015). Do ISO 9001-certified hotels get a higher customer rating thannon-certified ones? *International Journal of Hospitality Management, 51*(2015), 138–146.
Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD international conference on knowledge*. Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle; Washington, USA.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2016). *An introduction to statistical learning*, 6th Printing (p. 373). Springer.
Jones, P., Hillier, D., & Comfort, D. (2014). Sustainability in the global hotel industry. *International Journal of Contemporary Hospitality Management, 26*, 5e17.
Kandampully, J., & Suhartanto, D. (2000). Customer loyalty in the hotel industry: The role of customer satisfaction and image. *International Journal of Contemporary Hospitality Management, 12*(6), 346–351.
Kang, H., Yoo, S., & Han, D.. (2009). Accessing positive and negative online opinions. In Proceedings of the 13th international conference on human–computer interaction (HCII2009).
Kim, W. G., & Cha, Y. (2002). Antecedents and consequences of relationship quality in hotel industry. *International Journal of Hospitality Management, 21*(4), 321–338.
Lima, A. C. E. S., Castro, L. N. D., & Corchado, J. M. (2015). A polarity analysis framework for Twitter messages. *Applied Mathematics and Computation, 270*(2015), 756–767.
Litvin, S., Goldsmith, R., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468.
Liu, B., Hu, M. & Cheng, J.. (2005). Opinion Observer: Analyzing and comparing opinions on the web. Proceedings of the 14th international world wide web conference (WWW-2005), may 10-14, 2005, Chiba, Japan.
Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F., & Matsuo, Y. (2013). Identifying customer preferences about tourism products using an aspect-based opinion mining approach. 17th international conference in knowledge based and intelligent information and engineering systems - KES2013. *Procedia Computer Science, 22*(2013), 182–191.
Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management, 34*(1), 99–107.
McAlexander, J. H., Kaldenburg, D. O., & Koenig, H. F. (1994). Service quality measurement. *Marketing Health Services, 14*(3), 34.
Miao, Q., Li, Q., & Dai, R. (2009). Amazing: A sentiment mining and retrieval system. *Journal of Expert Systems with Applications, 36*, 7192–7198.
Myung, J., Lee, D., & Lee, S. (2008). A Korean product review analysis system using a semi-automatically constructed semantic dictionary. *Journal of KIISE: Software and Applications, 35*, 392–403.
Nasution, H. N., & Mavondo, F. T. (2008). Customer value in the hotel industry: What managers believe they deliver and what customer experience. *International Journal of Hospitality Management, 27*(2), 204–213.
NDTV (Nov, 2015). Available from: http://www.ndtv.com/goa-news/goa-tourism-to-go-online-for-tourists-convenience-1241119; Accessed 20.02.16.
Oliva, T. A., Oliver, R. L., & MacMillan, I. C. (1992). A catastrophe model for developing service satisfaction strategies. *The Journal of Marketing*, 83–95.
Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior, 31*, 527–541.
Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, pp. 1320–1326).
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). Association for Computational Linguistics.
Pang, B., Lee, L., & Vaithyanathan, S.. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the conference on empirical methods in natural language processing (EMNLP '02).
Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management, 55*(2016), 16–24.
Planet Goa (2016). Available from: http://planetgoaonline.com/; Accessed 10.06.16.
Radojevic, T., Stanisic, N., & Stanic, N. (2015). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism Management, 51*(2015), 13e21.
Rhee, H. T., & Yang, S. B. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior, 50*(2015), 576–587.
Rylander, R. G., Propst, D. B., & McMurtry, T. R. (1995). Nonresponse and recall biases in a survey of traveler spending. *Journal of Travel Research, 33*(4), 39–45.
Saif, H., He, Y., Fernandez, M., & Alani, H. (2015). Contextual semantics for sentiment analysis of twitter. Information processing and management. *Information Processing and Management, 52*(2016), 5–19.
Salehan, M., & Kim, D. J. (2015). Predicting the performance of online customer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems, 81*(2016), 30–40.
Sayeed, A. B., Graber, J. B., Rusk, B. & Weinberg, A.. (2012). Grammatical structures for word-level sentiment detection. Proceedings of the 2012 conference of the North American Chapter of the association for computational Linguistics: Human language technologies, Montreal, Canada, 2012.
Sista, S., & Srinivasan, S.. (2004). Polarized lexicon for review classification. In Proceedings of the international conference on machine learning, models, technologies and applications.
Sparks, B. A., So, K. K. F., & Bradley, G. L. (2015). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management, 53*(2016), 74e85.
Stine, R., & Foster, D. (2014). *Statistics for business decision making and analysis* (2nd ed., Vol. 343). Pearson New International Edition, 411, 593.
Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis Chinese documents. *Journal of Expert Systems with Applications, 34*, 2622–2629.
Thet, T. T., Na, J. C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 0165551510388123.
Torres, E. N., Singh, D., & Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. *International Journal of Hospitality Management, 50*(2015), 77–83.
Touropia (2015). Available from: http://www.touropia.com/tourist-attractions-in-india/; Accessed 24.02.16.
The Travel and Tourism Competitiveness Index Ranking (2015). Available from: http://reports.weforum.org/travel-and-tourism-competitiveness-report-2015/index-results-the-travel-tourism-competitiveness-index-ranking-2015/; Accessed 08.07.16.
Tripadvisor (2016). Available from: https://www.tripadvisor.in/Hotels-g297604-Goa-Hotels.html, Accessed on 08.01.16.
Tripzuki. (2013). *The latest trends in Indian tourism*. Available from: https://www.tripzuki.com/blog/the-latest-trends-in-indian-tourism/ Accessed 26.02.16.
Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In *Information, intelligence, systems and applications (IISA)* (pp. 1–6). IEEE, 2013 Fourth International Conference.
Trusov, M., Bucklin, R., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social media site. *Journal of Marketing, 73*, 90–102.
Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, 0894439310386557.
Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424). Association for Computational Linguistics.
Udemy Blog. (2014). *7 advantages of tourism in an economy*. Available from: https://blog.udemy.com/advantages-of-tourism/ Accessed 09.03.16.

UNEP. (2016). *Economic impacts of tourism*. Available from: http://www.unep.org/resourceefficiency/Home/Business/SectoralActivities/Tourism/WhyTourism/ImpactsofTourism/EconomicImpactsofTourism/tabid/78783/Default.aspx Accessed 09.03.16.
UNWTO. (2015). *United nations world tourism organization annual report 2015*.
Westbrook, R. (1987). Consumption-based affective responses and post-purchase processes. *Journal of Marketing Research, 24*(3), 258–270.
World Economic Forum. (2007). *World T&T competitiveness report*. Geneva.
Xie, K. L., Zhang, Z., & Hang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management, 43*, 1–12.
Xu, Y. (2010). *Status and future of hotel industry in China (in Chinese)*. Modern Business (p. 146e147).
Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *IJHM, 28*, 180–182.
Ye, Q., Zhang, Z., & Law, R. (2008). Sentiment classification of online reviews to travel destination by supervised machine learning approaches. *Journal of Expert Systems with Applications, 36*, 1–9.
Zhang, Y. (2011). *Stages of star hotel in China* (p. 62e65). Modern Hotel Research (in Chinese).
Zhang, Z., Ye, Q., Law, R., & Li, Y. (2011). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *IJHM, 29*(4), 694–700.
Zhou, L., Ye, S., Pearce, P. L., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management, 38*, 1–10.
Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2009). *Business research methods* (8th ed., p. 435).

Dr Geetha has done her Bachelors in Science and Masters in Business Administration from University of Madras; she has completed her Ph.D. in Marketing from Indian Institute of Technology Madras. She has around 11 years of teaching and research experience. Her research areas are impulse buying behaviour, variety seeking behaviour and store environment. She has published in journals such as European Journal of Marketing, Journal of Retailing and Consumer Services to name a few. She is a reviewer for many journals.

Pratap Singha and Sumedha Sinha are students, MBA program at Department of Management studies, Indian institute of Technology Madras.

Sumedha Sinha is student, MBA program at Department of Management studies, Indian institute of Technology Madras.