

CONTENT ANALYSIS OF ONLINE REVIEWS: A CASE STUDY OF TIMIȘOARA ACCOMODATION ESTABLISHMENTS

GORDAN MARIUS-IONUȚ*¹, ADAMOV TABITA¹, ȘUBA ANKA¹, IANCU TIBERIU¹

¹Banat's University of Agricultural Sciences and Veterinary Medicine "King Michael I of Romania" from Timisoara, Faculty of Management and Rural Tourism, Romania

*Corresponding author's e-mail: marius.gordan@gmail.com

***Abstract:** The aim of this research paper is to analyze online hotel reviews of accommodation establishments from Timișoara, Romania, written on the Tripadvisor and Booking.com platforms. The main goal of the research is to determine the main topics that tourists write about and their appreciation of attributes such as cleanliness, comfort, location, facilities, staff and value for money. The research paper also deals with analyzing the responses that owners wrote back to the reviewers, considering the preference for writing repetitive responses or customized ones and polite or impolite answers. Finally, we compare the overall distribution of positive and negative reviews between the two platforms, in order to pinpoint possible signs of astroturfing.*

Key words: reviews, tourism, content analysis, Timișoara, responses to reviews, astroturfing

INTRODUCTION

Reading online reviews has become one of the main tools that consumers have at their disposal in order to gain information about the products that they would like to purchase [8], superseding traditional marketing methods [6] in terms of reliability for the customer, understood as the correspondence between the projected image and product reality. Considering this, online reviews play a significant role in the decision-making process of the customers, with some sources that this form of electronic word of mouth is the single most important source of information [12].

Tourism has also been a major part of this trend, considering the emergence of dedicated websites hosting reviews for hotels and other accommodation establishments (e.g., Tripadvisor, Yelp) or online travel agencies having a dedicated section on their website attributed to reviews (e.g. Booking). In the context of decreasing information asymmetry between potential customers and hotel management and of the increase competitiveness between establishments, online hotel reviews can also be a useful tool for hotel managers in order to better understand what guests consider important during their stay [13]. All types of accommodation units can benefit from the competitive advantage provided by actively monitoring reviews and other kinds of user-generated content online, such as travel blogs, photographs and videos posted on Instagram or similar social media platforms [2]. Higher review scores are reported to improve income and profit rates for hotels and this should provide an incentive for managers to pay attention to the online image of their organization [15]. An emerging factor influencing the way that the public perceive tourism organizations are social media influencers [4]. In the context of the impact generated by their content, they differ from regular users in several ways, among them being the attention given to the production quality of text, photography or video content and the extended reach that this kind of content receives [10]. However, the platforms used by influencers differ from those used by users looking specifically for hotel reviews (Facebook, Twitter, YouTube, Tumblr, Tiktok [16], as opposed to Tripadvisor, Yelp, Booking, Google). While this does not mean that managers need to dismiss using influencers' services in the marketing plan of their organization, the immediate impact of the promotional campaign can become diffuse if the target audience of said influencer is different from that of the hotel or accommodation establishment. Considering this, online

reviews could play a larger role in the consumer decision making process of any type of consumer, not just followers of certain influencers.

Most platforms also provide the management team with the opportunity to respond openly to reviews [7]. Responses can become an effective customer relationship management tool [11] and at the same time a marketing technique, with both positive and negative consequences. At the same time, business decisions can be fine-tuned continuously based on the information received from customers [14], online reviews being a cost-efficient and non-intrusive way of generating valid data [9], as the reviews are willingly written and shared freely by their authors.

MATERIALS AND METHODS

Online reviews were obtained from the following websites: booking.com and tripadvisor.com, using a web scraper written in R. The following packages have been used in the data collecting stage: rvest [17], Rselenium [18], tidyverse, httr. Rvest was used for collecting data from both tripadvisor.com and booking.com, while booking.com required the usage of Rselenium, as the website contains JavaScript elements. Once the website was loaded using Rselenium, the actual data collection was done using rvest.

A total of 55.433 reviews were obtained from booking.com. After data cleanup 17.677 reviews were considered valid. Tripadvisor offered considerably fewer reviews with a total of 4.123 reviews scraped and only 1.776 valid instances.

Data cleanup has been done in Microsoft Excel and Google Sheets, by removing instances of reviews without usable text (text written in languages other than Romanian or English and reviews containing only the review score). In the interest of obtaining a larger database of reviews, automated translation of reviews written in other languages could be employed in the future. However, such methods were avoided in this study due to the fact that other studies report contradictory results regarding the usage of machine translation [1, 8]. Further developments in the field could alleviate this issue and grant easier access to the review data.

The following quantitative content analysis results have been obtained using WordStat: distribution of phrases for positive and negative segments of reviews, topic identification. Topics were grouped according to the attribute they describe: location, staff, price, cleanliness, facilities, comfort, food. Further analysis was done to determine the most commonly referred to attributes in the positive and the negative section of the reviews.

Qualitative content analysis was done to determine the way that the management team responds to reviews. Each accommodation unit that responded to tourists was assessed regarding the repetitiveness of the responses and the politeness shown. Responses that referred to the guests by their name or that didn't resemble other ones written by the same host were deemed customized responses, while the rest were considered repetitive. Responses that were accusatory and/or contained harsh words were deemed impolite.

Finally, in order to determine if some signs of astroturfing (dishonest reviews, commonly written by or on behalf of the owner in order to boost review scores, or conversely, negative reviews directed at rival organizations) [3] were present on the websites, we compared the means of the review scores between them using the independent sample t test (Welch's t test) [5]. The scores were standardized to a one to five scale. In order to reduce the chance of false positives, only accommodation establishments that have a high number of reviews were considered. The assumptions that make this analysis relevant are the following:

- Booking.com is less prone to astroturfing compared to Tripadvisor, as only verified guests can write reviews on the platform (only guests that made a

reservation at the accommodation establishment through the website). Therefore, astroturfing on booking.com would still be possible, albeit costly (each reservation received incurs a fee for the organization, a common amount being 15% deducted from total revenue).

- The organizations that decide to practice astroturfing would only do so extensively on one platform, given time and monetary constraints, with Tripadvisor being the primary target due to the lack of reservation required.

Further studies may be done in order to evaluate the validity of this approach.

RESEARCH RESULTS

As seen in figure 1 and 2, the most common phrases used in the positive sections of the reviews are related to the location attribute (city center, good location, close to the center, quiet area, excellent location). Guests also tend to mention positive experiences with the staff and/or host, with perceived level of comfort and available facilities being other topics that were common. Price would seem to play a relatively small role for the guests writing positive reviews when compared to other attributes, although it is still significant.



Figure 1: Word cloud of most common phrases in positive reviews

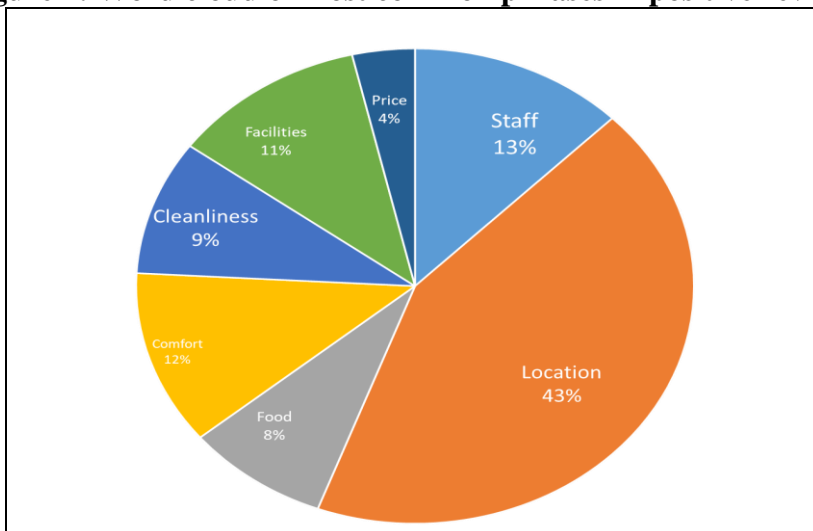


Figure 2: Attributes of most common phrases in positive reviews

As illustrated in figures 3 and 4, negative reviews mostly deal with the lack of thermal comfort (lack of air conditioning or central heating), water pressure and

temperature, inadequate parking spaces, poor breakfast, with cleanliness playing a relatively small, although still significant role. We note that the distribution of attributes between positive and negative reviews is unequal, as such we can infer that while some attributes are greatly appreciated when they are deemed positive traits by the guest (e.g. location), they tend to not even be acknowledged by guests when they write negative reviews. As such, other attributes are taken into consideration.

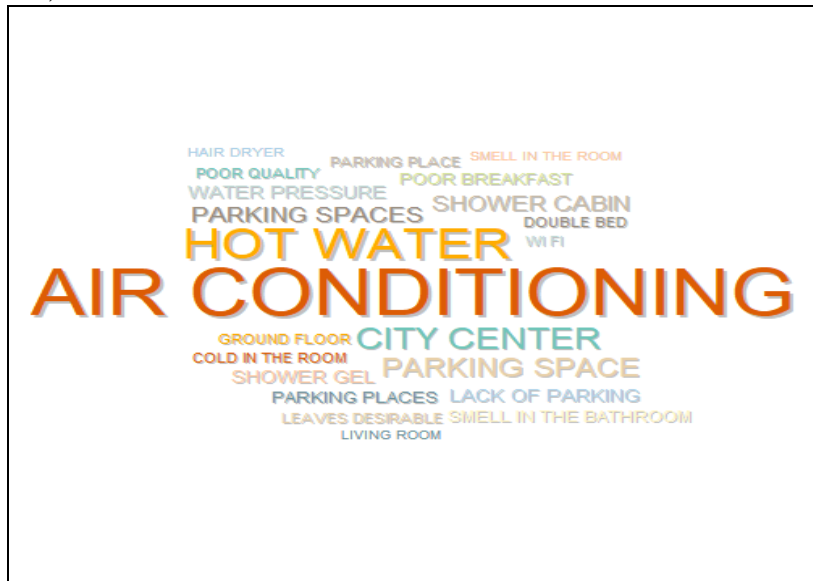


Figure 3: Word cloud of most common phrases in negative reviews

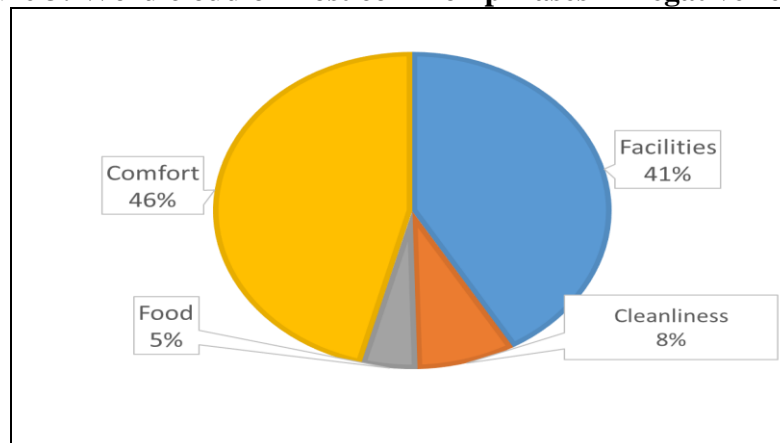


Figure 4: Attributes of most common phrases in negative reviews

86% of accommodation establishment owners or managers responded to reviews in a polite way, while in 81% of cases the hosts responded in a unique way to each guest. It should be noted that the pattern of responses indicates that almost all answers that are impolite are associated with customized, unique responses. This can be expected, considering that a repetitive response that would be appropriate to both positive and negative reviews is a sequence of thanks and possibly invitations to visit the accommodation establishment in the future, with the promise of improving conditions. The reverse is not true, as repetitive messages can be a sign of emotional detachment while writing responses, therefore impolite repetitive answers were not registered.

There are no distinct statistically significant variations between the review scores of the two sites, as shown in the table . In this case the research hypothesis coincides with the null hypothesis of the t-test, therefore a p-value over 0.05 is deemed appropriate to accept the research hypothesis. Only accommodation units that had a relatively large number of

reviews on both platforms were tested, as the power of the test decreases at lower sample values.

Table 1.**Welch's T-test applied to two independent samples of reviews**

Name of accommodation unit	Mean value of reviews on Booking	Mean value of reviews on Tripadvisor	P-value of t-test	Significant difference of means
Imperial Premium	3.82	3.59	0.109	Negative
NH Hotel	4.04	4.05	0.944	Negative
Hotel Timișoara	3.74	3.66	0.293	Negative
Hotel President	4.59	4.21	0.061	Negative
Hotel Silva	4.61	4.40	0.059	Negative*
Pensiunea Casa Leon	4.72	4.52	0.253	Negative
Hotel Casa del Sole	4.33	4.23	0.349	Negative
Freeborn Hostel	4.52	4.65	0.220	Negative
Hotel Ambassador	4.27	4.13	0.240	Negative
Hotel Iosefin Residence	4.39	4.40	0.970	Negative
Hotel Vanilla	4.73	4.58	0.081	Negative
Check Inn Hotel	4	4	1	Negative
Hotel Galaxy	4.56	4.40	0.171	Negative
Best Western Plus Lido Hotel	3.90	4.83	0.411	Negative

However, it was noted that some distinct signs of astroturfing were discovered during the qualitative phase of the content analysis, with three apartments owned by the same host having almost identical reviews. However, there were no indications that this is a systemic practice in Timișoara. More refined methods of detection may be employed in order to assess this phenomenon, using machine learning algorithms and a combination of text analysis and review score comparison between different websites.

CONCLUSIONS

In conclusion, analyzing hotel reviews can be an extremely useful tool, as they are available in a readily accessible data format on dedicated websites. R, Rstudio and the packages used for this research make web scraping for data a relatively easy and straightforward job, with minimal programming skills required for using functions and loops, in order to automate the process. Another advantage of this manner of data collection is the speed at which data is collected. If we do not take into consideration the time needed to write the program, thousands of reviews can be obtained and organized in minutes.

Most of the time the reviews are written in a concise manner, making data cleanup and processing relatively easy to do. At the same time, even before cleanup, the data obtained directly in R is formatted in an easy to parse way, as data frames can be written to a format that is compatible with other software environments (.csv can be exported to Excel, SPSS, Stata).

Understanding the way guests relate to certain attributes of the accommodation experience can be the first step that hotel owners and managers have to undertake in order to maximize the efficiency of their workload, by concentrating only (or mostly) on the few areas that guests care about. For example, providing guests with improved thermal comfort by properly insulating buildings, adding heating and cooling devices to hotel rooms or

investing in improved plumbing and water heating can be a way to reduce future complaints from guests.

On the other hand, given that location was the attribute corresponding to 43% of the most commonly used phrases in positive reviews, an investor might consider this sufficient proof of the importance of placing the building housing the accommodation establishment in a location that is most convenient to guests, such as near to the city center. This is not always possible, considering financial constraints, but the disadvantages derived from a non-central location can be overridden by investing in comfortable rooms and excellent facilities.

Review responses are another area that managers can pay attention to, as they can become a part of the customer relationship management routine in the organization. Special care should be taken to ensure that the response is written in a polite but empathetic manner, as those responses are public and can influence consumer behavior in the future. The vast majority of hosts that responded used polite language, which is a good sign regarding their professionalism. Continuous improvements in this field are required, as not all organizations can afford a specialized public relations team to manage this process, therefore extraordinary care should be directed to the way the responses are written by managers, with politeness, attention to detail and correct grammar being must-haves.

Finally, astroturfing acts as a distortion of both market mechanisms and data analysis, damaging customer trust in online reviews if the general scores and reviews are artificially inflated or deflated and do not represent reality and negatively influencing investment decisions. As such, using as many data sources as possible for analyzing specific hotels is imperative to guests, hosts and researchers. Difficulties may arise due to different website architectures, forcing the user to either manually collect data or adapt the automated data collection protocol. Even this approach is not faultless, as astroturfing can be done in a systemic manner, with detection methods mostly exploiting flaws in the review writing process.

REFERENCES

- [1]. **BALK E.M., CHUNG M., HADAR N.**, 2012, Accuracy of Data Extraction of Non-English Language Trials with Google Translate [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US)
- [2]. **CROTTS JOHN C., PEYTON R. MASON, BOYD DAVIS**, 2009, Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry: An Application of Stance-Shift Analysis to Travel Blog Narratives, *Journal of Travel Research*, pp. 139-151
- [3]. **CHO CHARLES., MARTENS MARTIN, KIM HAKKYUN ET AL.**, 2011, Astroturfing global warming: it isn't always greener on the other side of the fence. *J. Bus. Eth.* 104(4), pp.571–587
- [4]. **EUN KYOO KIM ELLEN, MATTILA ANNA, BALOGLU SEYHMUS**, 2011, Effects of Gender and Expertise on Consumers' Motivation to Read Online Hotel Reviews, *Cornell Hospitality Quarterly*, pp. 975-991
- [5]. **GRAEME D. RUXTON**, 2006, The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test, *Behavioral Ecology*, Volume 17, Issue 4, pp. 688-690
- [6]. **GRETZEL ULRIKE, KYUNG HYAN YOO**, 2008, Use and Impact of Online Travel Reviews, *Information and Communication Technologies in Tourism 2008* pp 35-46

- [7]. **GU BIN, YE QIANG**, 2014, First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction, Production and Operations Management, pp. 570-482
- [8]. **JEFFREY L. JACKSON, AKIRA KURIYAMA, ANDREEA ANTON et al.**, 2019, Accuracy of Google Translate for Abstracting Data From Non-English-Language Trials for Systematic Reviews. *Annals of Internal Medicine*, 171(9), pp.677
- [9]. **LI GANG, LAW ROB, HUY QUAN VU, JIA RONG, XINYUAN (ROY) ZHAO**, 2015, Identifying Emerging Hotel Preferences Using Emerging Pattern Mining technique, *Tourism Management*, pp. 311-321
- [10]. **MAGNOA FRANCESCA, CASSIA FABIO**, 2018, The impact of social media influencers in tourism, *Anatolia An International Journal of Tourism and Hospitality Research*, pp. 1-3
- [11]. **PAYNE ADRIAN, FROW PENNIE**, 2005. A strategic framework for customer relationship management. *J. Market.* 69 (4), pp.167-176
- [12]. **PHILLIPS PAUL, BARNES STUART, ZIGAN KRYSTIN, SCHEGG ROLAND** 2017, Understanding the Impact of Online Reviews on Hotel Performance: An Empirical Analysis, *Journal of Travel Research* , pp. 235-238
- [13]. **RIEGNER CATE**, 2007, Word of Mouth on the Web: The Impact of Web 2.0 on Consumer Purchase Decisions, *Journal of Advertising Research*, pp. 436-437
- [14]. **STUART LEVY, WENJING DUAN, SOYOUNG BOO.**, 2013, An Analysis of One-Star Online Reviews and Responses in the Washington, D.C., Lodging Market, *Cornell Hospitality Quarterly*, pp. 49-63
- [15]. **YE QIANG, ROB LAW, BIN GU, WEI CHEN**, 2011, The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of E-word-of-mouth to Hotel Online Bookings, *Computers in Human Behavior* 27 (2), pp. 634-636
- [16] **ZHANG TENGHAO, HUANG XINLI**, 2020, Viral marketing: influencer marketing pivots in tourism-a case study of meme influencer instigated travel interest surge, *Current Issues in Tourism*, pp. 1-8
- [17]. ***<https://cran.r-project.org/web/packages/rvest/rvest.pdf>
- [18]. ***<https://cran.r-project.org/web/packages/R Selenium/RSelenium.pdf>