Tourism Management 57 (2016) 80–83

Contents lists available at ScienceDirect

Tourism Management

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

Research note

Effects of the Booking.com scoring system

Juan Pedro Mellinas*, Soledad-María Martínez María-Dolores, Juan Jesús Bernal García

Department of Quantitative Methods and Computing (Faculty of Business Studies), Technical University of Cartagena (UPCT), Calle Real, 3, CP. 30201, Cartagena, Murcia, Spain

HIGHLIGHTS

- A special scoring system used by Booking leads to awarding better scores than would using a conventional system.
- Differences are clearly more significant in hotels with low scores than in those with high scores.
- On hotels with very high scores, Booking gets worse scores.

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form
31 March 2016
Accepted 25 May 2016

Keywords:
Hotel
Reviews
Booking.com
Scores

ABSTRACT

Booking.com provides a large travellers' reviews database that is useful for consumers, hoteliers and academics. Recent research discovered unexpected peculiarities in its scoring system, suggesting it could lead to "inflated scores" for hotels on their website. To confirm the above suspicions, the methodology used in this paper compares values assigned on Booking.com with those on Priceline that have a conventional scoring system. Conclusions show significant distortions, particularly in hotels with low and medium scores.

1. Introduction and background

The phenomenon of online reviews of products and services is similar to traditional "word-of-mouth" (WOM). The importance of this type of marketing as economic activity has been described in the academic literature (Duan, Gu, & Whinston, 2008; Goldenberg, Libai, & Muller, 2001; Stokes & Lomax, 2002; Zhu & Zhang, 2010). When this system is adapted to the Internet by means of an organized system of reviews and scores, its importance multiplies and new opportunities open up for all participants (Dellarocas, 2003; Pan, MacLaurin, & Crotts, 2007). That is why interaction with these processes has become an essential component in the marketing strategy development of many organizations (Chen & Xie 2005). The spread of WOM through the Internet and Web 2.0 is currently known as "electronic Word of Mouth" (eWOM) (Gruen, Osmonbekov, & Czaplewski, 2006).

The tourism sector and, more specifically, the hotel sector are especially affected by eWOM. Millions of travellers write reviews about their travel experiences, seemingly receiving nothing in return. This altruistic phenomenon is described by Resnik and Zeckhauser (2002) and is considered an online reproduction of similar phenomena of friendliness and collaboration that exist in the offline or "real world". This passion for sharing experiences and opinions has created large databases that allow rapid and convenient access to both business and academics, who extract relevant information about customer satisfaction in the hotel industry.

Several studies show that online reviews written by travellers influence consumer decisions (Gretzel & Yoo, 2008; Vermeulen & Seegers, 2009). People believe that these reviews are more objective than commercial information (Ricci & Wietsma, 2006; Blackshaw & Nazarro, 2006). Guests are willing to pay higher prices to hotels with a good record (Yacouel & Fleischer, 2012; Melnik & Alm, 2002).

Some authors have chosen Booking.com (Booking) as a source of information for research. Most of these studies assume that Booking works with a 0–10 or 1–10 scale, but the minimum score for hotels on this website is 2.5 (Bjørkelund et al., 2012). A recent

* Corresponding author.
E-mail addresses: losmellinas@yahoo.es (J.P. Mellinas), soledad.martinez@upct.es (S.-M. Martínez María-Dolores), juanjesus.bernal@upct.es (J.J. Bernal García).

http://dx.doi.org/10.1016/j.tourman.2016.05.015
0261-5177/

paper explained this anomaly, clearly describing how the Booking scoring system actually works (Mellinas, María-Dolores, & García, 2015) and showing how it assigns 2.5 to the worst reviews. This implements a 2.5–10 scale instead of the expected 0–10 or 1–10 scale. The authors suggested that it could be associated with "inflated scores" on this website, affecting customer perceptions and academic research results.

We conduct a study to confirm this hypothesis and determine whether the distortion affects all types of hotels. Scores obtained from Booking and Priceline.com (Priceline) for a sample of hotels are compared because Priceline uses a conventional 1–10 scale. Using this methodology, we can determine whether both systems provide similar scores or whether there are significant variations.

2. Booking and Priceline

Although we can perceive two very different companies, Priceline actually acquired Booking in 2005. Priceline is one of the world's leading providers, with others brands such as agoda.com, Kayak, rentalcars.com, and OpenTable. The Priceline corporation has different brand strategies, similar to what happens with Expedia.com and its Hotels.com division. Brian Ek, a spokesman for Priceline, said, "Booking is the world's top hotel-reservation service, while Priceline is the leading provider of discount-travel services in the USA" (Levy, 2013).

Despite the minor marketing differences, results when searching for a hotel are quite similar. Regarding valuation systems, both show an apparent 0–10 or 1–10 traditional scale. In the case of Booking, the rating system is not as expected (Mellinas et al., 2015) but presents special features. The final score published is a plain average of six aspects of the hotel. Furthermore, the customer is not asked to rate each item from 1 to 10; instead, there are four options corresponding to poor (2.5), fair (5), good (7.5) or excellent (10).

To check the Priceline scoring system features, we made a reservation and completed the survey after our stay. We were invited to evaluate every aspect of the hotel using a scale from 1 to 10. Furthermore, we had to evaluate the "Overall Experience", which is the score published with every single review. The Priceline system is what we would expect from a system based on a 1–10 scale; therefore, it is a valid reference for comparison with Booking.

Observing both sites' reviews, we realized that Booking deletes reviews older than 14 months, whereas Priceline keeps them up to 24 months. This policy contributes to an always-updated information database for both sites, as opposed to other websites, such as TripAdvisor, that retain reviews for years.

3. Methodology and sample

The unexpected features of the Booking scoring system may lead to "inflated" scores (Mellinas et al., 2015), but quantitative data which can confirm this are not provided. We consider that the best way to check if scores are really inflated is to use a sample of hotels and see if the valuations obtained from Booking are really higher than those obtained from other websites. When searching among the most popular websites (Travelocity, Orbitz, Expedia, Hotels and TripAdvisor), we observed that all of them show a system with a maximum score of 5 (apparently a scale 1–5).

In our opinion, performing score transformations to enable a comparison is not appropriate. Transforming data from these other sites (multiplying by two) would result in a 2–10 scale; transforming Booking scores (dividing by two) would result in a 1.25–5 scale. Either approach might cause slight distortions, which are unacceptable when seeking accurate results.

Therefore, we opted for the 1–10 scale Priceline system, whose performance is exactly that expected by most researchers who have used Booking as a database in the past (Chaves, Gomes, & Pedron, 2012; Costantino, Martinelli, & Petrocchi, 2012a, 2012b; Filieri & McLeay, 2014; Korfiatis & Poulos, 2013; Martínez et al., 2012; Plata-Alf, 2013; Yacouel & Fleischer, 2012). This methodology allows us to establish accurate comparisons between actual values on Booking and those that would be obtained from a system with a real 1–10 scale.

We tested if mean difference in both systems is 0 (H_0), in order to corroborate what was suggested in the cited article, i.e., that the Booking.com approach to scaling review scores is associated with "inflated scores", and to quantify that distortion. We did not just make a comparison of the mean scores on both sites but instead did a more elaborate analysis, which allowed us to obtain additional information. We also studied whether differences are more significant in lower rated hotels.

Although Priceline offers hotel reservations around the world, as indicated above, it is mainly focused on the United States market. For this reason, we decided to use hotels located in this country for our sample design, to have access to hotels with a large number of reviews. We set the condition that the selected hotels must have at least 100 reviews recorded on both websites. Due to the aforementioned policy of removing old reviews, it was not easy to find hotels for a large sample. We divided the sample design in four stages:

a) We made a selection of major cities and tourist destinations, where it is more likely to find a significant number of large hotels, which met the conditions set.
b)We selected those destinations with at least 5 valid hotels.
c)We identified the number of valid hotels and classified the cities in three groups:
- 20 or more hotels.
- 10 to 19 hotels.
- 5 to 9 hotels.
d) In the first group we only found Las Vegas, with 53 hotels and randomly selected 50. In the second group (10 destinations) we randomly selected 10 hotels in each destination and 5 hotels for the third group (10 destinations).

Data were collected during the first week of May 2014 and final selected sample includes 200 hotels in 21 different locations. Fifty of the hotels are located in Las Vegas, 10 locations have 10 hotels (San Diego, Los Angeles, San Francisco, Chicago, Atlanta, Orlando, New York, Atlantic City, Boston and Washington DC), and another 10 destinations have 5 hotels (New Orleans, Philadelphia, Dallas, Houston, Phoenix, Miami, Memphis, Baltimore, Long Beach and Honolulu). The fact that Las Vegas has 50 hotels in the sample is justified by the number of large hotels (25 of the 50 largest hotels in the world) located in this popular tourist destination.

We collected 67,946 reviews from Priceline and 130,306 from Booking. In the case of Priceline, all but two hotels have more than 200 reviews. However, in the case of Booking, only 141 exceed this figure. On the other hand, some hotels have over 1000 reviews; Booking has 38 of these, but Priceline has just one.

4. Results

Compared scores show higher ratings on Booking in most of the sample, with an average score of 7.99; the average rating on Priceline is 7.52. Moreover, 78% of hotels have higher scores on Booking, 5% have an identical score on both websites, and only 17% have higher scores on Priceline.

We designed a graphical representation of results for each hotel, reflecting values obtained on Priceline on the x-axis and values obtained on Booking on the y-axis. A cluster of dots was obtained

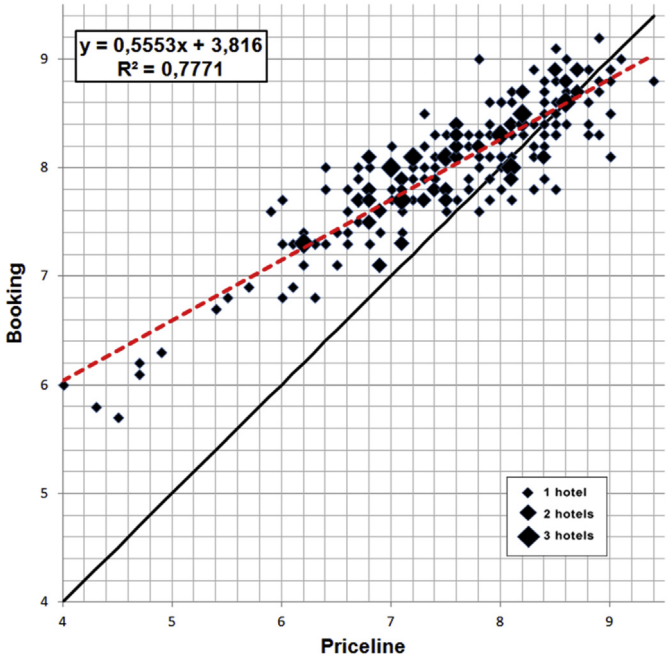

Fig. 1. Scores on priceline and booking.
Source: Analysis using data collected from Priceline and Booking

(Fig. 1); a bigger dot represented cases where the scores of two or three hotels matched. This cluster of dots suggested the presence of a significant linear correlation; therefore, we calculated the corresponding linear regression, represented by the red dashed line. Also we draw a line dividing the chart into two equal parts, which should be all dots if the scores on Priceline and Booking were identical. Dots above this line represent cases of hotels with better scores on Booking and dots below this line hotels with better scores on Priceline.

The R-square value of 77.7% confirmed a significant relationship between rating level and differences between the two websites. Note that differences in scores are more remarkable for hotels with lower valuations, becoming smaller as those valuations increase. Actually, hotels with very high valuation obtain better scores on Priceline. For hotels with valuations up to eight points on Priceline, scores are clearly better on Booking, representing almost two thirds of the sample (64.5%). There is also a group of hotels, with scores ranging from 8.1 to 8.5 (22.5% of the sample), without a website providing clearly higher scores. Finally, the group with scores above 8.6 (only 13% of the sample) shows hotels with scores predominantly better on Priceline.

Based on these data, we rejected the null hypothesis (H_0) about the mean equality of scores in Booking and Priceline, but we found that the effect of "inflated scores" is not equal for all hotels, depending on the rating level. We can say that there are very significant differences (in favour of Booking) for hotels with low scores, small differences in average-rated hotels and an absence of differences for hotels with high scores. Only in the case of hotels with very high scores do we observe a weak superiority of scores on Priceline.

In Fig. 2, we represent a frequency distribution of values in both sites. We note that Booking scores are highly concentrated above the score of 7.3, as ratings below this figure only accounted for 8% of the sample, while 36.5% are in that range for Priceline. The Booking scoring system seems to minimize the number of hotels with low scores, moving them to a range with medium scores.

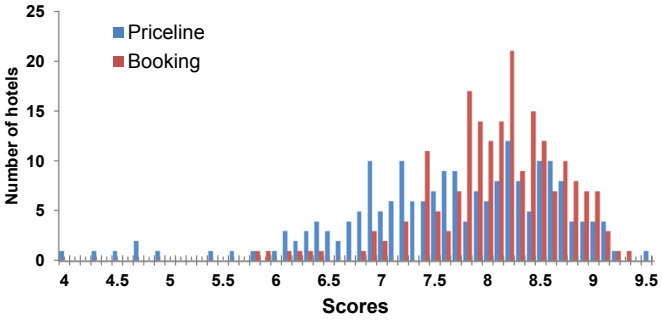

Fig. 2. Frequency distribution for Priceline and Booking reviews.
Source: Analysis using data collected from Priceline and Booking

5. Conclusions

Data supported suspicions of "inflated scores" derived from the Booking scoring system. Thus, our hypothesis is confirmed, but the results provide two additional conclusions:

- Differences between both scoring systems are clearly more significant in hotels with low scores than in those with high scores.
- On hotels with very high scores, the Booking system does not contribute to inflating the scores but results in lower scores than a conventional system such as Priceline.

The first conclusion was certainly expected because one of the Booking system peculiarities is the 2.5 minimum score for every review. Because the score of each hotel is the arithmetic mean of six different aspects, if some of them are not rated with the minimum score (2.5), the resulting score is even above 2.5.

The second conclusion seems more surprising, but we suspect it can be explained by the average of six aspects, which does not stimulate very high scores. To receive a maximum score of 10, it is necessary that the client assign the maximum score for each of the six aspects; otherwise, the average would be reduced. Conversely, in the case of Priceline, a very satisfied customer can give a 9 or 10, although has not rated all aspects with the highest score.

Findings shown in this paper should contribute to reconsidering numerical analysis when using data provided by Booking. Future research could be oriented toward avoiding the arithmetic mean that the system performs automatically, studying every single aspect in an individual way.

This analysis has methodological limitations related to the sample because, although the size and quality may be appropriate, it is limited to a single country and the comparison is made with a single website. It would be interesting to perform similar studies with different samples, selecting hotels from other countries and other websites that use a 1–10 or 0–10 scoring system. Moreover, additional quantitative research about possible effects on sales, that these differences in scores between websites may induce, should be conducted in the future.

Acknowledgements

This research was partially supported by the Faculty of Business Studies at the Technical University of Cartagena (UPCT)

References

Bjørkelund, E., Burnett, T. H., & Nørvag, K. (2012). A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference*

on information integration and web-based applications & services (pp. 229–238). ACM.
Blackshaw, P., & Nazzaro, M. (2006). *Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer. A Nielsen BuzzMetrics white paper* (2nd ed.). Spring.
Chaves, M. S., Gomes, R., & Pedron, C. (2012). Analysing reviews in the Web 2.0: small and medium hotels in Portugal. *Tourism Management, 33*(5), 1286–1287.
Chen, Y., & Xie, J. (2005). Third-party product review and firm marketing strategy. *Marketing Science, 24*(2), 218–240.
Costantino, G., Martinelli, F., & Petrocchi, M. (2012a). Priorities-based review computation. In *AAAI spring symposium: Intelligent web services meet social computing*.
Costantino, G., Morisset, C., & Petrocchi, M. (2012b). Subjective review-based reputation. In *Proceedings of the 27th annual ACM symposium on applied computing* (pp. 2029–2034). ACM.
Dellarocas, C. (2003). The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407–1424.
Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing, 84*(2), 233–242.
Filieri, R., & McLeay, F. (2014). E-WOM and accommodation an analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research, 53*(1), 44–57.
Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Marketing Letters, 12*(3), 211–223.
Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and Communication Technologies in Tourism*, 35–46.
Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). eWOM: the impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business research, 59*(4), 449–456.
Korfiatis, N., & Poulos, M. (2013). Using online consumer reviews as a source for demographic recommendations: a case study using online travel reviews. *Expert Systems with Applications, 40*(14), 5507–5515.
Levy, A. (2013). *Booking.com challenging parent priceline in U.S. Travel (Bloomberg). Disponible en.* http://www.bloomberg.com/news/2013-01-22/booking-com-challenging-parent-priceline-in-u-s-travel.html.
Martínez María-Dolores, S. M., Bernal García, J. J., & Mellinas, J. P. (2012). Los hoteles de la región de Murcia ante las redes sociales y la reputación online. *Revista de Análisis Turístico*, (13), 1–10.
Mellinas, J. P., María-Dolores, S. M. M., & García, J. J. B. (2015). Booking. com: the unexpected scoring system. *Tourism Management, 49*, 72–74.
Melnik, M. I., & Alm, J. (2002). Does a seller's ecommerce reputation matter? Evidence from eBay auctions. *The Journal of Industrial Economics, 50*(3), 337–349.
Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research, 46*(1), 35–45.
Plata-Alf, D. (2013). Marketing communications in a virtual environment – Opportunities and challenges for companies in the tourism sector. In A. Nalepka, & A. Ujwary-Gil (Eds.), *Business and non-profit organizations facing increased competition and growing customers' demands*. Wyższa Szkoła Biznesu-National-Louis University.
Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: empirical analysis of eBay's reputation system. *Advances in Applied Microeconomics, 11*, 127–157.
Ricci, F., & Wietsma, R. T. (2006). Product reviews in travel decision making. *Information and Communication Technologies in Tourism*, 296–307.
Stokes, D., & Lomax, W. (2002). Taking control of word of mouth marketing: the case of an entrepreneurial hotelier. *Journal of Small Business and Enterprise Development, 9*(4), 349–357.
Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management, 30*(1), 123–127.
Yacouel, N., & Fleischer, A. (2012). The role of cybermediaries in reputation building and price premiums in the online hotel market. *Journal of Travel Research, 51*(2), 219–226.
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *Journal of Marketing, 74*(2), 133–148.

Juan Pedro Mellinas is researcher at Department of Quantitative Methods and Computing (Faculty of Business Studies), Technical University of Cartagena (UPCT), Cartagena, Spain. Degree in administration and management (Universidad de Murcia). Master in Tourism (UPCT). Ph.D in administration and management.

Soledad María Martínez María-Dolores is Professor at Department of Quantitative Methods and Computing (Faculty of Business Studies), Technical University of Cartagena (UPCT), Cartagena, Spain. Degree in Business and economy. Ph.D in administration and management.

Juan Jesús Bernal García is Professor at Department of Quantitative Methods and Computing (Faculty of Business Studies), Technical University of Cartagena (UPCT), Cartagena, Spain. Director of Department of Quantitative Methods and Computing.