University of Reading

Effects of the booking.com rating system: bringing hotel class into the picture

Article

Accepted Version

Mariani, M. M. and Borghi, M. (2018) Effects of the booking.com rating system: bringing hotel class into the picture. Tourism Management, 66. pp. 47-52. ISSN 0261-5177 doi: https://doi.org/10.1016/j.tourman.2017.11.006 Available at http://centaur.reading.ac.uk/74560/

Effects of the Booking.com rating system:
Bringing hotel class into the picture

(Marcello Mariani, Henley Business School and Matteo Borghi)

Abstract

The purpose of this study is to continue the discussion initiated by Mellinas et al. (2015, 2016) on the effects of the Booking.com rating system and more widely on the use of the OTA as a data source in academic tourism and hospitality research. We enrich and complement the original work by Mellinas et al. (2015) by empirically investigating the effects of the Booking.com rating system on the distribution of hotel ratings for the overall population of hotels located in London over two years. Based on more than 1.2 million online reviews, we show that the overall distribution of hotel scores is significantly left-skewed. Moreover, we find that the degree of skewness is positively associated with hotel class: lower-class hotels exhibit distributions of ratings that are statistically less skewed than higher-class hotels.

Key words: Booking.com, online reviews, London, big data, hotel ratings, hotel class, skewness.

1. Introduction

Researchers in tourism and hospitality are addressing a significant amount of research questions related to electronic word-of-mouth (eWOM) (Cantallops & Salvi; 2014). They have collected and analysed online consumer reviews (OCRs) published on the most popular independent consumer review website TripAdvisor (Fang et al., 2016; Xie et al., 2014), Online Travel Agencies (OTAs) such as Expedia (e.g., Stringam & Gerdes, 2010) and Booking (e.g., Mellinas et al., 2016). Furthermore, they have compared multiple online review websites (Xiang et al., 2017).

However, to date scarce attention has been paid to issues problematizing critically the use of OCRs websites as a data source (Yacouel & Fleischer, 2012; Mellinas et al., 2015, 2016). In this work, we tackle some of these issues and add to previous methodological findings related to the use of Booking.com hotel reviews (Mellinas et al., 2015) by showing that not only the distribution of hotel ratings is left skewed but also that the characteristics of the distribution change according to the hotel class. This research might allow researchers in hospitality and tourism to critically assess and contextualise their and others researchers' findings when Booking.com online reviews are employed as data source. Accordingly, this work is needed and timely as it addresses the problematic lack of disclosure and explanation of how online ratings are distributed, and contributes to enhance the level of methodological rigour when applying quantitative methods to data sourced from such OTAs as Booking.com.

2. Online Travel Agencies and electronic Word Of Mouth in tourism and hospitality research

Online reviews have become an object of study in the marketing management literature since the nineties when the topic of eWOM has been described by leveraging a relationship marketing approach in Internet contexts (Stauss, 1997; Hennig Thurau et al., 2004).

The development and consolidation of OTAs have brought about a proliferation of reviews on hospitality and tourism services to date. The aforementioned trends have triggered an unprecedented development of research on the topic of eWOM in hospitality and tourism (Gretzel & Yoo, 2008; Vermeulen & Seegers, 2009; King et al., 2014). Building on the approach introduced in the systematic literature review carried out by Cantallops and Salvi (2014), studies in hospitality have dealt with review-generating factors (e.g., Kim et al., 2009; Nusair et al., 2011) and on the impacts of eWOM from a consumer perspective (e.g., Wen, 2009; Sparks and Browning, 2011) and a company perspective (e.g., Dickinger, 2011; Loureiro & Kastenholz, 2011; Ye et al., 2009).

Despite the plethora of studies published recently by leveraging OCRs to understand the generating factors and impact of eWOM (Banerjee & Chua, 2016; Fang et al., 2016; Filieri, 2016), the methodological issues related to the use of OCRs websites as a data source in academic hospitality and tourism research have been largely neglected (Yacouel & Fleischer, 2012; Mellinas et al., 2015, 2016).

This issue is particularly relevant when considering the OTA Booking.com. Indeed, it apparently holds the largest share of certified hotel reviews worldwide (Revinate, 2017). Despite an increased scholarly appetite for analyses comparing multiple platforms (Xiang et al., 2017), most of extant literature has relied on the OTA Expedia (e.g., Stringam & Gerdes, 2010; Xiang et al., 2015; Xiang et al., 2017). One exception is the study of Mellinas et al (2015), showing that the actual rating scale used by Booking is a 2.5 – 10 scale and warnng that assuming that Booking deploys a 0-10 or 1-10 scale might generate statistical inaccuracies if data are used for research. The authors suggest that Booking apparently inflates scores to the higher end of the scale, but do not provide an empirical substantiation of the actual distribution of reviews. In their ensuing study, Mellinas et al. (2016) find that differences in scores are more significant in hotels with low scores than in those with high scores and that Booking gets worse scores for hotels with very high scores.

However, neither of the two studies by Mellinas et al. (2015, 2016) illustrates the distributions of the entire populations of hotels in the selected destinations, thus constraining the generalizability of results related to the OTA's scale and scoring systems. Secondly, the distributions analysed are never characterised in terms of skewness and kurtosis and therefore we lack empirical evidence that the

Booking.com rating distributions analysed are statistically and significantly skewed for large samples. Third, there is no specific analysis of the factors associated with the inflated scores at the hotel level.

To address the aforementioned issues, we first reconstruct the distribution of Booking.com online ratings for the entire population of hotels located in London (UK) to assess and increase the generalizability of extant research (Mellinas et al., 2015). Second, we compare the distribution of hotel reviews based on the hotel class to understand if the distribution differs across class. Accordingly, our study is distinctive for three reasons as it: 1) builds on the overall population of Booking OCRs over a two years period in London; 2) examines the distributions of hotel online ratings based on their skewness and kurtosis; 3) compares the distributions of online ratings by hotel class and tests if the aforementioned distributions are statistically equal. Overall, we provide a methodological contribution to improve researchers' knowledge of OCRs websites as a data source for their quantitative empirical analyses.

3. Methodology

Building on a data science approach (Witten et al.., 2016) to tourism management (Geetha et al., 2017; Xiang et al., 2015), we quantitatively examined the distribution of hotel ratings after retrieving and analysing large amounts of data (Berthold et al., 2011).

We retrieved the overall population of Booking.com OCRs for hotels based in London (UK) over a two years time frame (from January 5, 2015 to January 5, 2017). Since our objective was to examine quantitatively if the scores are somewhat inflated towards the higher end of the scale for an entire population of hotels, we focused on London as it ranks third among the Top 100 city destinations worldwide (Euromonitor International, 2017).

The data retrieval and analysis were performed by using a web crawler that was developed in Python programming language. We obtained a total of 1,228,089 reviews across 866 hotels, with 862 out of 866 hotels exhibiting at least one review. Data was analysed by leveraging a number of data science techniques and tools, including Data Understanding (DU) (Berthold et al., 2011; Witten et al.., 2016). To study the characteristics of the overall distribution of ratings (and the distributions by hotel class), we deployed nonparametric kernel density estimators. We used the Tn-statistic proposed by Li et al. (2009) and tests reported in Maasoumi and Racine (2002) to test if the sub-distributions (one for each hotel class) were equal.

4. Findings

The distributions of the overall populations of hotel ratings in London is presented below (see Fig. 1) together with the box-plot (see Fig. 2).

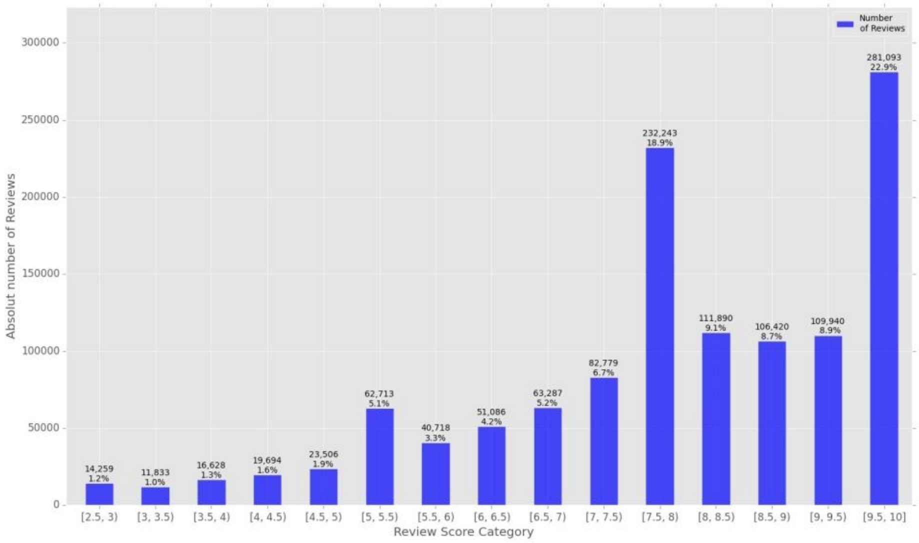

Figure 1. Distribution of Booking.com hotel reviews' scores, London, 5 Jan 2015 – 5 Jan 2017.

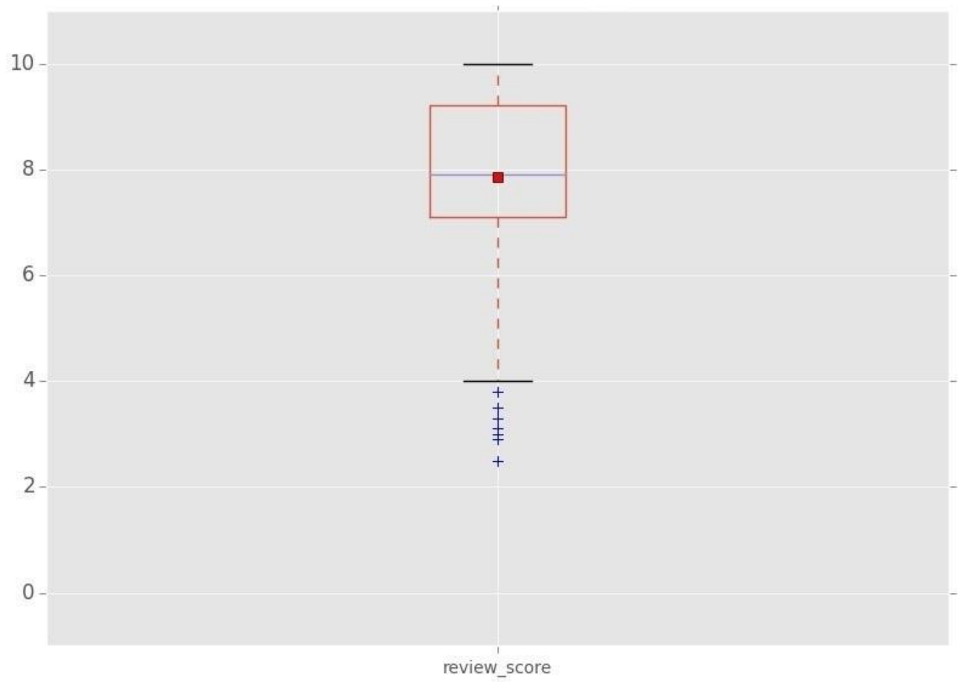

Figure 2. Box-plot of Booking.com hotel reviews' scores, London, 5 Jan 2015 – 5 Jan 2017.

Figure 1 shows that most of hotel review scores (68.5%) are concentrated in an interval between 7.5 and 10. The boxplot of Fig.2 further confirms that average and median values are close to 8.0. More specifically, 50% of the scores are included within the interval 7.1-9.2.

Given that the actual rating scale used by Booking is a 2.5 – 10 scale and not a 0-10, and since the actual survey is based on a "smileys" scale with four different options, most of the hotel ratings tend to cluster in practice around the following values: 5.0, 7.5 and 10.0. The overall distribution is left-skewed (skewness = -0.8135729). Apparently, the skewness of the distribution is dependent on the peculiar scoring system deployed by Booking.com.

Interestingly, if we split the overall population of hotel ratings across four subsamples based on hotel class, we come up with the following distributions (see Fig. 3):

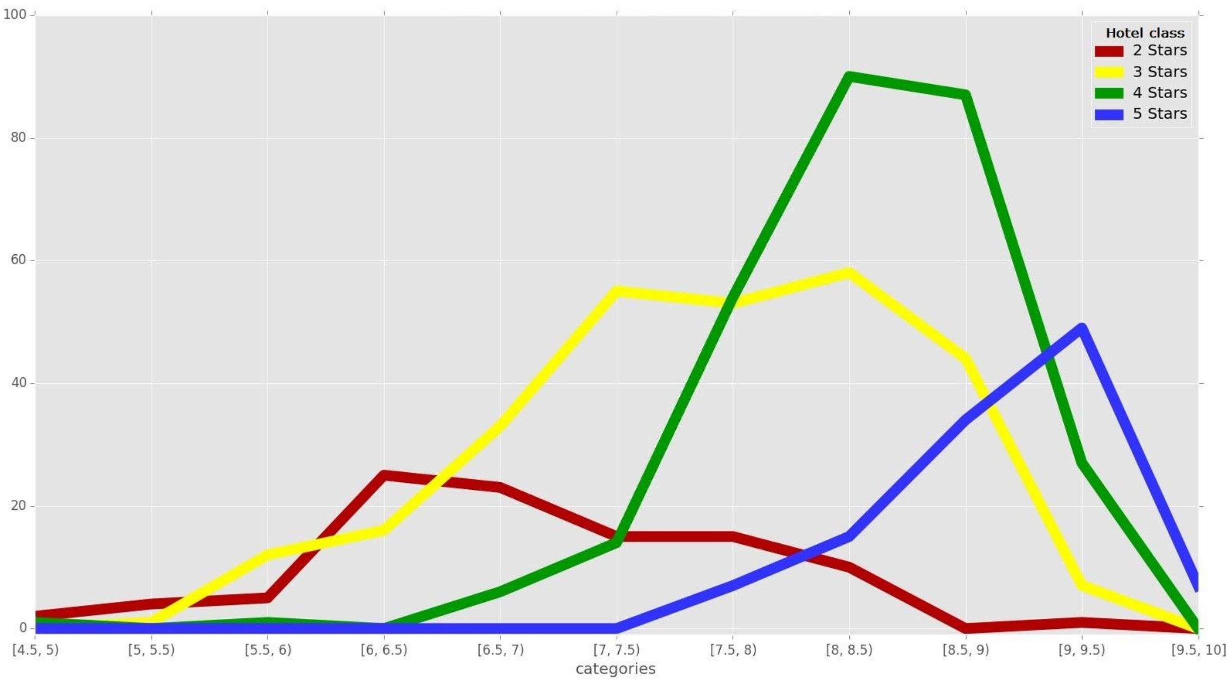

Figure 3. Distribution of Booking.com hotel reviews' scores, based on hotel class subsamples

The characteristics of the four distributions are reported below (see Table 1):

	Two-star	**Three-star**	**Four-star**	**Five-star**
Number of Hotels	128,407	450,684	469,854	91,097
Skewness	-0.3247548	-0.6981048	-1.012907	-1.640329
Kurtosis	2.412263	3.080939	3.661301	5.764365

Table 1. Skewness and kurtosis for the four distributions of Booking.com hotel reviews' scores

Table 1 shows that the distributions of scores for all of the four sub-samples are left-skewed. Moreover, it seems that the higher the hotel class the more the distribution is left-skewed. That said, the average and median values of the distributions of the four subsamples tend to increase as we move from a lower to a higher hotel class (indeed mean and median values are equal to 6.8 and 7.1 respectively for 2-star hotels; 7.7 and 7.9 for 3-star hotels; 8.2 and 8.3 for 4-star hotels; 8.8 and 9.2 for 5-star hotels).

To study the characteristics of the four distributions and their differences, we deployed nonparametric kernel density estimators (see Fig. 4).

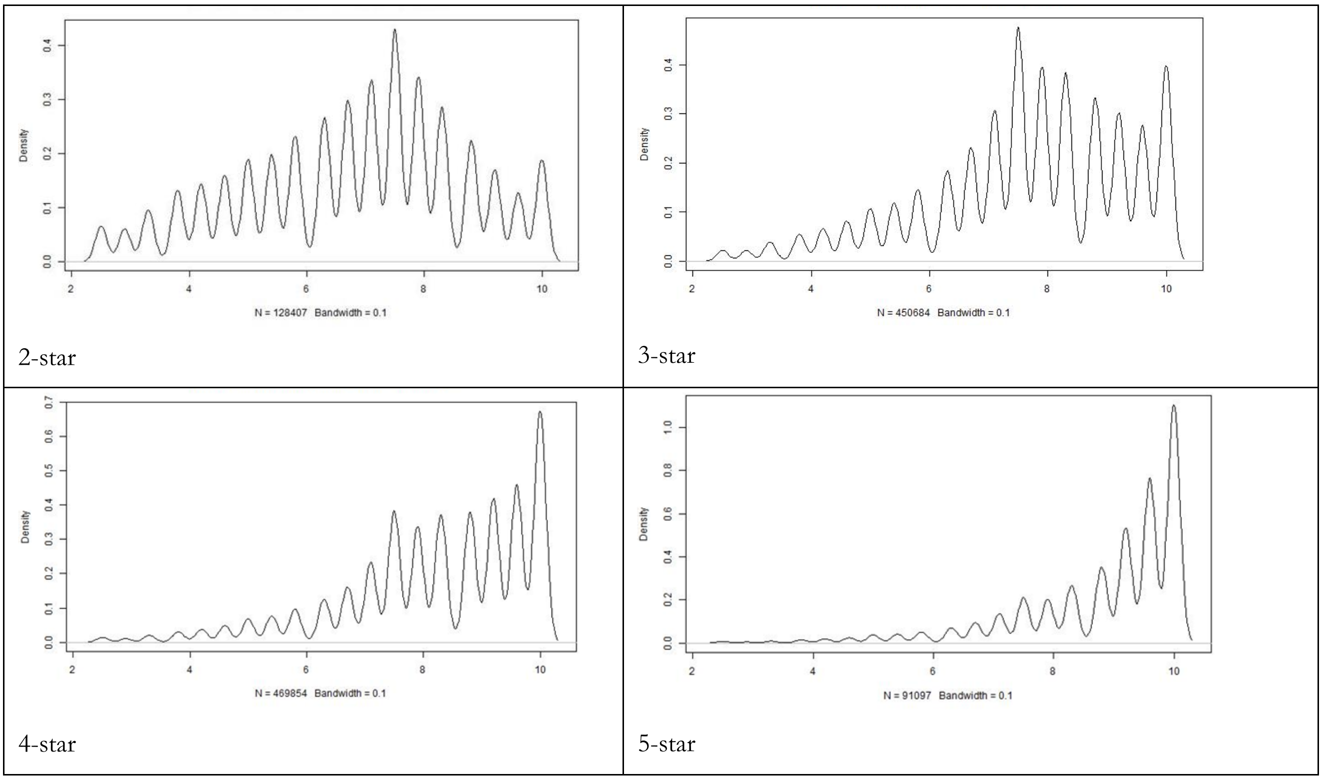

Figure 4. Kernel density distributions

In particular, we have used the Tn-statistic proposed by Li et al. (2009) and tests reported in Maasoumi and Racine (2002) to test if the four distributions are equal by comparing them in pairs, with our null hypothesis being the equality. The null hypothesis of equality is rejected at the 0.1% level in all the comparisons. We therefore conclude that all the distributions, despite being left-skewed, are statistically different. Furthermore, the higher the hotel class, the more left-skewed is the distribution of ratings. This significantly complements the results obtained by Mellinas et al. (2015, 2016) and might improve the accuracy of findings of researchers deploying Booking as a data source.

5. Conclusions and implications

This study has contributed to a thin yet emerging research line pertaining to issues related to the use of OTAs as data sources in academic hospitality and tourism research (Mellinas et al., 2015, 2016). First, we empirically substantiate that the distribution of review scores on Booking.com is left-skewed due to the scoring system used by Booking (Mellinas et al. 2015). This evidence emerges when analysing the ratings of the entire hotel population of London and is consistent with what seems to emerge implicitly from other platforms (e.g., Tripadvisor and Expedia) when looking at sentiment scores and their correlation with online ratings (Xiang et al., 2017). Second, we found that the non-normality of ratings' distributions is associated with the hotel class: more specifically, the higher the hotel class, the more skewed are the distributions of ratings. Third, we argue that using Booking.com ratings in statistical models assuming normality of data might require preliminary ad hoc transformations bringing the distributions closer to a normal distribution.

Relevant methodological implications stem from this study. First, scholars using online ratings stemming from Booking.com should disclose their distributions (not meeting the criteria of normality) and ensure they do not use statistical models assuming normality of data. Second, they should carry out robustness checks to understand if and to what extent the distribution differences across hotel class depend on structural factors (i.e., the scoring system) or factors related to the product or service reviewed (i.e., the attributes and features of the product/service reviewed) and assess how this might affect their findings.

This research is not without limitations. While our findings stem from an empirical analysis of the overall population of hotels in a specific destination (London), they could possibly vary across destinations. Therefore, further empirical research might be needed to test if the findings can be generalised to other destinations. Moreover, more analyses should be carried out to disentangle the scoring system effect and the hotel class effect. In other terms it would be interesting to investigate if it is the Booking.com scoring system that affects differently the distributions of ratings across hotel class or rather (as we expect) the hotel class per se is a driving factor for differentiated skewness because guests of high class hotels are systematically more satisfied with their accommodation.

REFERENCES

Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, 53, 125-131.

Berthold, M.R., Borgelt, C., Hoppner, F., Klawonn, F. & Ada, I. (2011). *Guide to Intelligent Data Analysis*. London: Springer Verlag.

Cantallops, A. S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, 41–51.

Dickinger, A. (2011). The trustworthiness of online channels for experience- and goal-directed search tasks. *Journal of Travel Research*, 50 (4), 378–391.

Euromonitor International (2017). Top 100 City Destinations Ranking. Retrieved on the 22nd of June 2017 from http://go.euromonitor.com/rs/805-KOK-719/images/2017%20Top%20100%20Cities%20Destinations%20Final%20Report.pdf

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

Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research* , 58, 46-64.

Geetha, M., Singha, P. & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. *Tourism Management*, 61, 43-54.

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and Communication Technologies in Tourism*, 35-46.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1), 38–52.

Kim, T., Kim, W.G., Kim, H.B. (2009). The effects of perceived justice on recovery satisfaction, trust, word-of-mouth, and revisit intention in upscale hotels. *Tourism Management*, 30, 51–62.
King, R.A., Racherla, P., Bush, V.D. (2014). What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature. *Journal of Interactive Marketing*, 28(3), 167-183.
Li, Q., Maasoumi, E. and Racine, J.S. (2009), A nonparametric test for equality of distributions with mixed categorical and continuous data, *Journal of Econometrics*, 148(2), pp. 186-200.
Loureiro, S.M.C. & Kastenholz, E. (2011). Corporate reputation, satisfaction, delight, and loyalty towards rural lodging units in Portugal. *International Journal of Hospitality Management* (30), 575–583.
Maasoumi, E. & Racine, J.S. (2002). Entropy and predictability of stock market returns, *Journal of Econometrics*, 107(1/2), 291-312.
Mellinas, J. P., María-Dolores, S. M. M., & García, J. J. B. (2016). Effects of the Booking.com scoring system. *Tourism Management*, 57, 80-83.
Mellinas, J. P., María-Dolores, S. M. M., & García, J. J. B. (2015). Booking.com: The unexpected scoring system. *Tourism Management*, 49, 72-74.
Nusair, K., Parsa, H.G., Cobanoglu, C. (2011). Building a model of commitment for Generation Y: an empirical study on e-travel retailers. *Tourism Management*, 32, 833–843.
Revinate (2017), *Global Hotel Reputation Benchmark Report 2017*, accessed on the 30/6/2017 from https://learn.revinate.com/hospitality-research-studies/global-hotel-reputation-benchmark-report-2017 .
Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310-1323.
Stauss, B. (1997). Global Word of Mouth. Service Bashing on the Internet is a Thorny Issue. *Marketing Management*, 6(3), 28 –30.
Stringam, B.B. & Gerdes, J. Jr. (2010). An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management*, 19(7), 773-796.
Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123-127.
Wen, I. (2009). Factors affecting the online travel buying decision: a review. *International Journal of Contemporary Hospitality Management*, 21 (6), 752–765.
Witten, I.H., Frank, E. Hall, M.A. & Pal, C.J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Fourth Edition. Morgan Kaufmann: Cambridge, MA, USA.
Xiang, Z., Du, Q., Ma, Y., Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65.
Xiang, Z., Schwartz, Z., Gerdes, J., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44(1), 120-130.
Xie, K., Zhang, Z. & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1-12.
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.
Ye, Q., Law, R., Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* 28, 180–182.