DOES VERIFYING USES INFLUENCE RANKINGS?
ANALYZING BOOKING.COM AND TRIPADVISOR

EVA MARTIN-FUENTES,* CARLES MATEU,† AND CESAR FERNANDEZ†

*Department of Business Administration, University of Lleida. Lleida, Spain
†INSPIRES Research Institut, University of Lleida, Lleida, Spain

Electronic word of mouth (eWOM) is of recent and considerable importance in tourism, particularly because of the intangible nature of the industry. Users’ online reviews are a source of information for other consumers, who take them into account before making a reservation at a lodging property. The aim of this study is to establish whether or not the anonymity of the reviews on TripAdvisor alters hotel rankings by comparing them with verified users’ reviews on Booking.com. Moreover, the study analyzes whether or not the differences in the rating scales of both websites favor some hotels over others. A large amount of data is used in this study, with more than 40,000 hotels on Booking.com and 70,000 on TripAdvisor in 447 cities around the world, comparing the rankings of about 20,000 hotels matched on both websites. Our findings suggest that the behavior of both rankings is similar and the lack of veracity on TripAdvisor due to the anonymity in the user’s verification system is baseless. In addition, some differences are found depending on the hotel category and region, due mainly to the unique rating scale on Booking.com (from 2.5 to 10) compared with the rating scale on TripAdvisor (from 1 to 5).

Key words: Electronic word of mouth (eWOM); TripAdvisor; Booking.com; Ranking; Rating scale

Introduction

Since its creation, TripAdvisor, a travel information website whose content is provided mostly by users, has had its share of controversy and has been widely questioned to the point of going to court over allegations by managers from hospitality establishments that feel harmed by the site’s content (Grindlinger, 2012). The main criticism of this platform is the perceived lack of veracity (Palmer, 2013; Rawlinson, 2011; Smith, 2012). Hoteliers complain about the anonymity of TripAdvisor; they argue the site allows anonymous users to give opinions about any establishment without having stayed there or used the establishment (Webb, 2014).

Conversely, Booking.com is a hotel-booking website that allows comments to be made by customers.
The site claims that it publishes “verified reviews from real people” because leaving a review on this website is only possible if an individual books an accommodation through the site and actually stays at the reviewed property.

These two websites are significant in the tourism sector and to the online reputation of the accommodation facilities reviewed. They are also important in terms of the research done before making a reservation—the percentage of which continues to rise (C. K. Anderson, 2012)—because consumer reviews generate more confidence than communications from the company (Gretzel & Yoo, 2008; Vermeulen & Seegers, 2009).

Moreover, being number one in a ranking is something to which every business aspires because potential customers see the top positions first (Spink & Jansen, 2006). A higher position results in a decision to use either tourism products (Ghose, Ipeirotis, & Li, 2012) or any other type of product (Pope, 2009; Sorensen, 2007), which therefore leads to more bookings (Ye, Law, Gu, & Chen, 2011). According to Filieri and McLeay (2013), product ranking emerges as the strongest antecedent of high-involvement travelers’ adoption of information from online reviews.

The aim of this study is to establish whether or not the anonymity of the reviews on TripAdvisor alters hotel rankings by comparing them with the verified users’ reviews on Booking.com. Alternatively, a study about the effects of the Booking.com scoring system confirms suspicions of “inflated scores” derived from their scoring system (Mellinas, Martínez María-Dolores, & Bernal García, 2016).

Taking into account that the scoring scale on Booking.com is from 2.5 to 10 and on TripAdvisor from 1 to 5—keeping in mind the research gap Mellinas et al. (2016) presented about the interest in performing similar studies with different samples and selecting hotels from other countries—the study seeks to compare whether or not the differences in the scales favor some hotels over others by their category and location.

Although eWOM for tourism is widely studied, the validity of anonymous reviews receives little attention. This article tries to fill this gap and helps incorporate research with the validity of reviews posted without verification procedures. Moreover, this article tries to provide additional insight into the effects of the Booking.com scoring system, complementing that of Mellinas et al. (2016) by analyzing a large volume of data from hotels worldwide and not just from a single country, comparing scores on “sale websites” (Booking.com) and “advice websites” (TripAdvisor) (Fernández-Barcala, González-Díaz, & Prieto-Rodríguez, 2010) and comparing the behavior of the scales according to the hotel categories.

Theoretical Background and Study Hypotheses

Word of Mouth

The word of mouth (WOM) phenomenon is studied in marketing (Arndt, 1967) and refers to client communications relating to a consumer experience (E. W. Anderson, 1998). With the advent of the internet, the way in which WOM reviews are made has been extended thanks to consumer-opinion portals (COPs) (Burton & Khammash, 2010), which allow consumers to review products and services and other people to view these online reviews and contribute to attenuating the negative effects of asymmetric information (Martin-Fuentes, 2016). WOM, propagated via Web 2.0, is known as “electronic word of mouth” (eWOM) (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004) and is defined as “all informal communications directed at consumers through Internet-based technology related to the usage or characteristics of particular goods and services, or their sellers” (Litvin, Goldsmith, & Pan, 2008, p. 461), and it has implications for tourism marketing (Morosan, 2013).

Recent eWOM studies have been conducted in relation to goods and services (Cheung & Thadani, 2012; Chevalier & Mayzlin, 2006). According to Cantallos and Salvi (2014), those focusing on the hotel industry can be split into review-generating factors (previous factors that encourage consumers to write reviews) and eWOM impacts (impacts caused by online reviews) from the point of view of consumers and companies.

Research based on 50 articles about hospitality and tourism eWOM concludes that “online reviews appear to be a strategic tool that plays an important role in hospitality and tourism management, especially in promotion, online sales, and reputation management” (Schuckert, Liu, & Law, 2015, p. 618).
TripAdvisor is one of the most influential eWOM sources in the context of hospitality and tourism (Yen & Tang, 2015), and it is often a hotel manager’s first point of call because of the significance the site has acquired for any accommodation facility’s reputation (Xie, Zhang, & Zhang, 2014). TripAdvisor is ranked as the most reliable source according to the perceptions of general managers (Torres, Adler, & Behnke, 2014). Numerous studies have been based on data provided by TripAdvisor (Ayeh, Au, & Law, 2013; Liu, Pennington-Gray, Donohoe, & Omodior, 2015; Mayzlin, Dover, & Chevalier, 2012; Melian-Gonzalez, Bulchand-Gidumal, & Gonzalez Lopez-Valcarcel, 2013; O’Connor, 2008; Vermeulen & Seegers, 2009). The percentage of consumers who consult TripAdvisor before booking a hotel room continues to increase (C. K. Anderson, 2012), and online rating lists are more useful and credible when published by well-known online travel communities like TripAdvisor (Casaló, Flavián, Guinalíu, & Ekinci, 2015).

In TripAdvisor’s own words, “the world’s largest travel site, enables travelers to unleash the full potential of every trip.” TripAdvisor branded sites make up the largest travel community in the world, reaching 350 million unique monthly visitors, with more than 570 million reviews and opinions covering more than 7.3 million accommodations, airlines, attractions, and restaurants (TripAdvisor, 2017a). However, TripAdvisor’s fame is also accompanied by some criticism because of the opportunity to comment anonymously without needing to have enjoyed the services of the hospitality industry (Gerrard, 2012); cases of a lack of truthfulness have been reported by journalists (Palmer, 2013; Rawlinson, 2011; Smith, 2012), as have cases of blackmail (Morrison, 2012; Webb, 2014). For example, some customers have tried to force establishment owners to offer a discount or invite them stay another day by telling owners they would leave negative comments on social websites if their demands were not met.

Booking.com is an online accommodation-booking website where travelers can compare prices and customer reviews (Neirotti, Raguseo, & Paolucci, 2016). The site claims to have 1,586,740 properties in 228 countries and to deal with over 1.5 million room-night reservations per day (Booking.com, 2017). Travel intermediary websites such as Booking.com are a popular online source for hotel information, as are social media websites like hotel information, as are social media websites like TripAdvisor and Facebook (Sun, Fong, Law & Luk, 2015), which draw the attention of researchers (Balagué, Martin-Fuentes, & Gómez, 2016; Viglia, Minazzi, & Buhalis, 2016).

Anonymity

It is possible for user-generated content to be anonymous because during the registration process on COPs, identities can be invented. This anonymity leads some reviewers to question the authenticity of the ratings posted on TripAdvisor (O’Connor, 2008) or discount their credibility (Duan, Gu, & Whinston, 2008). Moreover, anonymous users can post fake reviews to increase the reputation of their businesses or to harm their competitors (Dellarocas, 2003). Consumers often perceive eWOM as less trustworthy than WOM because it is difficult to identify the issuer, because user-generated content (UGC) is often created anonymously (Sparks & Browning, 2011; Yoo & Gretzel, 2010, cited by Leung, Law, van Hoof, & Buhalis, 2013).

As explained previously, to post a review on TripAdvisor there is no need to demonstrate that a user has actually stayed at a hotel; thus, the biggest threat to TripAdvisor is a loss of credibility (O’Connor, 2008). This fact is not the case with Booking.com because after booking a room through the website and staying at the accommodation the customer receives an invitation via e-mail to write a comment about the experience. So Booking.com only publishes reviews from users who have booked at least 1 night at a lodging property through its website and then stayed there. They therefore guarantee that the opinion is real —of which the system is aware.

Rankings

Rankings are used to quickly compare different options, thereby reducing the information time. If we consider that travelers economize on time and effort when they search for information (Solomon, Russell-Bennett, & Previte, 2012, cited by Filieri & McLeay, 2013) then rankings are a good way to obtain information on any kind of product or service.
Different studies show that rankings can have a significant impact on the decision to purchase a product or service. Among the different studies, worthy of note include Sorensen (2007) on the impact of the *New York Times* Best Sellers List on sales, that of Jin and Whalley, (2007) on how rankings affect the financial resources of public colleges, and that of Pope (2009) on the U.S. News & World Report ranking and its importance in the choice of hospitals in the US.

As confirmed by Filieri and McLeay (2013), product ranking is now the strongest antecedent of high-involvement travelers’ adoption of information from online reviews.

Taking into account that most users’ internet search engines only look at the first three pages of results (Spink & Jansen, 2006), that search results on Google decrease depending on their position in the ranking (Chitika, 2013), and that users rarely read reviews beyond the first web page (Pavlou & Dimoka, 2006), rankings are a highly valuable source of information and ranking in the top positions can draw the attention of potential customers.

Therefore, we propose the following hypothesis:

**H1**: Hotel rankings on TripAdvisor and Booking.com are related although they have different verification systems.

**Rating Scales on Booking.com and TripAdvisor**

Because of the importance of Booking.com and TripAdvisor to the hospitality sector, there are many studies using these two websites based on UGC for hotel experiences through content analysis techniques (Barreda & Bilgihan, 2013), or the analysis of reviews on each site with 50 hotels from Portugal (Chaves, Gomes, & Pedron, 2012).

The hotel scoring scale used by Booking.com is from 2.5 to 10, as shown in Figure 1 and as confirmed by Mellinas, Martínez María-Dolores, and Bernal García, (2015). The scoring scale used by TripAdvisor is from 1 to 5, rounding decimals to the midpoint.

Mellinas et al. (2015) conclude that this scale explains why most hotels have ratings above 7 and point out a research gap in the necessity to study the effects of this scoring system by quantifying how it inflates values.

Bjørkelund, Burnett, and Nørvåg (2012) compare hotels on TripAdvisor and on Booking.com. From a set of more than 600,000 reviews on Booking.com, they find that none are lower than 2.5 and conclude that Booking.com scores lean toward the higher end of the scale more often than TripAdvisor.

Mellinas et al. (2016) compare hotel scores between Booking.com and Priceline and asserted that there is no mean equality of scores, and that there are very significant differences (in favor of Booking.com) for hotels with low scores, low differences in average rated hotels, and an absence of differences for hotels with high scores. Hotels with a very high scores on Priceline have a weak superiority of scores.

Taking into account the limited literature on the unique rating scale of Booking.com, the following hypothesis is proposed:

**H2**: The rating scale on Booking.com (2.5–10) generates the same hotel scores as the scale on TripAdvisor (1–5).

**Methodology**

In this study we analyze the hotels of the top destinations in the world according to the 2015 TripAdvisor Ranking, dividing them into four regions, as proposed by Banerjee and Chua, (2016): America (AME), Asia and Pacific (ASP), Europe (EUR), and the Middle East and Africa (MEA). We then split these regions into countries and cities.

In April 2016 we automatically gathered the rankings of the hotels on Booking.com and TripAdvisor: the number of reviews on both websites, the ranking and scoring, hotel name, city and country, and the hotel category (the latter of these variables is the hotel star category according to Booking.com).

The data were collected using an automatically controlled web browser (developed in Python) that simulates a user navigation (clicks and selections) for TripAdvisor and Booking.com. Once the data were available, a new data set was created by joining corresponding data for a given hotel from both websites. The joint criteria were used for every city if:

- the hotel name was exactly the same;
- the hotel name from one site was contained, entirely, in the name from the other site (the choosing of
two lists (69,997 hotels on TripAdvisor and 40,580 on Booking.com), we automatically compared the hotels listed on both websites. The result was 20,880 hotels that matched both websites. The missing values were eliminated from all variables and the final result was 19,660 hotels, as shown in Table 1.

There were 11,871,134 reviews on Booking.com and 8,812,826 on TripAdvisor. To calculate the score (and therefore the position in the ranking), Booking.com does not take into account hotels with fewer than 5 reviews, so the minimum number of reviews on this website was 5, whereas on TripAdvisor it was 1, as shown in Table 1. The review mean is higher on Booking.com than on TripAdvisor.

The data collection from each destination was conducted simultaneously from both sites in order to have minimum variation. Because both websites are active and the data are modified over time, the data were extracted in less than 48 hr.

![Figure 1](image.png) Form sent by Booking.com to rate accommodation.
in order to determine the strength of the correlation. Missing values are eliminated from both variables to obtain identical pairs.

In relation to H2, we denote As and Bs as the standardized values of scores for each hotel. As the data sets for TripAdvisor and Booking.com have different rating scales (from 1 to 5 for TripAdvisor and from 2.5 to 10 for Booking.com), both data sets were standardized to a scale from 0 to 1. Furthermore, samples were paired by score for the same hotel on both websites. Thus, H2 null and alternative hypotheses could be stated as:

\[ H2_0 : \overline{as} - \overline{bs} = 0 \]

\[ H2_1 : \overline{as} - \overline{bs} \neq 0 \]

where \( \overline{as} \) and \( \overline{bs} \) are the mean values of As and Bs, respectively. In this case, the test statistic is:

\[
\frac{(tss - rss)}{(p - 1) \cdot rss} \cdot \frac{n}{n(p - 1)}
\]

following a Student’s t distribution with \( n - 1 \) degrees of freedom (df).

Because we worked with a large volume of data and applied the central limit theorem, the population of sample means was assumed to be normal.

Results

For H1 with the hotels that match on both websites, the Spearman correlation coefficient is \( r_s = 0.87 \) and the p-value for the test statistic (Eq. 2) is \( p < 0.001 \). Thus, evidence is found to support the relationship between the two rankings. Table 2 shows the results after analyzing the data by region (AME, ASP, EUR, and MEA), subregion (Africa, Asia, Caribbean, Central America, Europe, Middle East, North America, South America, and South Pacific), and city.

As shown in Table 2, by subregion, the strongest correlations are in Europe and the Middle East, and the weakest are in the Caribbean because there are only 68 hotels spread across 9 countries.

Moreover, the analysis by city with the highest number of hotels (Paris, Istanbul, Rome, London, Bangkok, and Tokyo) shows that, apart from Istanbul, they have a strong Spearman correlation coefficient.
dotted lines define the confidence interval for predicted observations, and the black line plots the fitted model. In Figure 2 all of the aggregate data are plotted and in Figure 3 plots refer to regions.

As explained, Booking.com uses a scale from 2.5 to 10 and TripAdvisor from 1 to 5. With these scales, the minimum ratings for the hotels is 3 on Booking.com and 1 on TripAdvisor, meaning that among more than 20,000 hotels the minimum on Booking.com is never reached. However, the maximum ceiling of the scales (10 and 5, respectively) is reached in both cases. Booking.com calculates the global score with a mean of six items (value, services, comfort, clean, staff, and location). On TripAdvisor, the user gives a direct score and, later on, can assess other sections individually. However, these are not taken into account in the final score. From among the items rated by users, the one with the highest mean is staff, followed by location and comfort, as shown in Table 3.

The countries with the lowest scoring hotels on TripAdvisor are in Malaysia, the Dominican Republic, Singapore, Denmark, and Indonesia; on Booking.com they are in Egypt, China, India, the Dominican Republic, and Malaysia. The highest rated hotels on TripAdvisor are in Bermuda, Fiji. In general, and except for certain cities with few hotels, the results show a direct positive relationship between rankings. Thus, there is some relationship between the Booking.com and TripAdvisor rankings because, in most cases, a strong statistically significant correlation is obtained, indicating that the position in the two rankings is similar.

As with the previous tests, the regressions are statistically significant at $p < 0.001$ in all regions (AME, ASP, EUR, and MEA), so evidence is found to support the idea that there is a relationship between the rankings of TripAdvisor and Booking.com.

With all datasets, regression analyses establish a statistically significant relationship between the two variables ($F = 28050; \beta = 0.60, p < 0.001$). Thus, as the TripAdvisor position increases, the Booking.com position is likely to increase, and vice versa. Therefore, there is a statistically significant relationship between both rankings, as can be seen in Table 2.

Figure 2 and Figure 3 plot the fitted model represented in Equation 1. The gray lines plot the confidence interval for the mean observed values, the

<table>
<thead>
<tr>
<th>Subregion/City</th>
<th>n</th>
<th>$r_s$</th>
<th>$\beta$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AME</td>
<td>4,130</td>
<td>0.85**</td>
<td>0.83**</td>
<td>1.46**</td>
</tr>
<tr>
<td>ASP</td>
<td>5,922</td>
<td>0.83**</td>
<td>0.21**</td>
<td>2,513.00**</td>
</tr>
<tr>
<td>EUR</td>
<td>8,809</td>
<td>0.90**</td>
<td>0.86**</td>
<td>36,710.00**</td>
</tr>
<tr>
<td>MEA</td>
<td>799</td>
<td>0.83**</td>
<td>0.45**</td>
<td>1,968.00**</td>
</tr>
<tr>
<td>Africa</td>
<td>289</td>
<td>0.76**</td>
<td>0.16**</td>
<td>157.60**</td>
</tr>
<tr>
<td>Asia</td>
<td>5,573</td>
<td>0.82**</td>
<td>0.20**</td>
<td>2,256.00**</td>
</tr>
<tr>
<td>Caribbean</td>
<td>68</td>
<td>0.67**</td>
<td>0.23**</td>
<td>20.37**</td>
</tr>
<tr>
<td>Central America</td>
<td>167</td>
<td>0.80**</td>
<td>0.81**</td>
<td>455.10**</td>
</tr>
<tr>
<td>Europe</td>
<td>8,809</td>
<td>0.90**</td>
<td>0.86**</td>
<td>36,710.00**</td>
</tr>
<tr>
<td>Middle East</td>
<td>510</td>
<td>0.90**</td>
<td>0.55**</td>
<td>3,940.00**</td>
</tr>
<tr>
<td>North America</td>
<td>2,254</td>
<td>0.83**</td>
<td>0.79**</td>
<td>6,401.00**</td>
</tr>
<tr>
<td>South America</td>
<td>1,641</td>
<td>0.88**</td>
<td>0.88**</td>
<td>7,998.00**</td>
</tr>
<tr>
<td>South Pacific</td>
<td>349</td>
<td>0.79**</td>
<td>0.74**</td>
<td>627.10**</td>
</tr>
<tr>
<td>Paris</td>
<td>855</td>
<td>0.85**</td>
<td>0.75**</td>
<td>2,324.00**</td>
</tr>
<tr>
<td>Istanbul</td>
<td>607</td>
<td>0.70**</td>
<td>0.94**</td>
<td>571.80**</td>
</tr>
<tr>
<td>Rome</td>
<td>524</td>
<td>0.86**</td>
<td>0.63**</td>
<td>1,319.00**</td>
</tr>
<tr>
<td>London</td>
<td>485</td>
<td>0.90**</td>
<td>0.66**</td>
<td>1,935.00**</td>
</tr>
<tr>
<td>Bangkok</td>
<td>337</td>
<td>0.86**</td>
<td>0.76**</td>
<td>883.60**</td>
</tr>
<tr>
<td>Tokyo</td>
<td>331</td>
<td>0.73**</td>
<td>0.51**</td>
<td>369.00**</td>
</tr>
<tr>
<td>Shanghai</td>
<td>290</td>
<td>0.76**</td>
<td>0.11**</td>
<td>127.20**</td>
</tr>
<tr>
<td>Berlin</td>
<td>285</td>
<td>0.81**</td>
<td>0.67**</td>
<td>458.20**</td>
</tr>
<tr>
<td>Montreal</td>
<td>66</td>
<td>0.96**</td>
<td>0.72**</td>
<td>620.80**</td>
</tr>
<tr>
<td>Dublin</td>
<td>74</td>
<td>0.95**</td>
<td>0.63**</td>
<td>610.90**</td>
</tr>
<tr>
<td>Hamburg</td>
<td>155</td>
<td>0.30**</td>
<td>0.23**</td>
<td>11.88**</td>
</tr>
</tbody>
</table>

**$p < 0.001$.**
Anguilla, and Nicaragua; on Booking.com they are in Bermuda, Fiji, Guatemala, and Honduras.

Bearing in mind the scoring scale and rating method of each website, it is possible to observe that most reviews are in the upper half of the scale, so both systems are systematically positively biased all over the world. Thus, on TripAdvisor, 95.68% of the hotels are rated at 3 or more (midpoint on the scale), whereas on Booking.com, 95.57% of the hotels are rated above 6.25 (midpoint on the scale). On TripAdvisor, the results are similar in all regions, but on Booking.com we can see that in MEA and ASP the percentage above the mean is lower than in AME and EUR.

At the top of the scale, we find that 3.47% of the hotels on TripAdvisor obtain the maximum score, while only 1.29% of the hotels on Booking.com do the same. On average, the hotels that achieve the highest score are those in AME on TripAdvisor (3.93), as well as on Booking.com (8.11), and the hotels with the highest dispersion are those in MEA (0.67) (Tables 4–6).

Figure 3. Linear model between TripAdvisor and Booking.com ranking by regions.
To check H2 with the standardized scale (0–1) for both websites, a Student’s t test was done with pairs of variables, as the hotels analyzed are the same in the two different rankings.

The results allowed us to reject the null hypothesis that the score means are equal on both websites, so we confirmed that on TripAdvisor the score mean is lower than on Booking.com \( \left[ t(19,659) = -4.86, p < 0.001 \right] \). However, the effect size is negligible (Cohen’s \( d = 0.03 \)) with all datasets, so an in-depth analysis split by regions and by hotel category was carried out. Table 7 shows the Student’s t test by region and by hotel category in which the rating scale mean is higher for five-star hotels worldwide on TripAdvisor, is statistically significant, and has the largest effect size of all datasets, especially, the ones from MEA.

On Booking.com, the scoring scale benefits one- to three-star hotels in AME and EUR with a low-medium effect size, and is detrimental to four-star hotels in ASP and in MEA with a low effect size. For all other hotels, the null hypothesis is accepted, as the mean differences are equal and/or the effect size is negligible.

Discussion

For most of the cities analyzed, the results show that there is a high degree of relationship between both websites’ rankings, indicating very strong statistically significant correlations. They likewise show that the possible publication of fake reviews on TripAdvisor do not seem to be prevalent, as both rankings behave similarly. Hence, having analyzed the data, it can be said that the verification systems of both websites do not affect the position of hotels.

The rankings of both websites were checked and the fact that they present such a high statistically significant correlation is important given their usefulness for comparing different options, thereby reducing the time and effort needed to identify the most suitable accommodation (Filieri & McLeay, 2013).

With regard to the second hypothesis, the unique rating scale of Booking.com (from 2.5 to 10) compared to TripAdvisor’s scale (from 1 to 5) benefits one- to three-star hotels in AME and EUR and is detrimental to five-star hotels worldwide, and four-star hotels in ASP and MEA. The reason why five-star hotels have higher scores on TripAdvisor than on Booking.com could be that the score is assigned directly, whereas on Booking.com the scoring system is the arithmetic average of six elements, as shown in Figure 1 (comfort, value, clean, staff, location and services). Thus, it is necessary for all users rating a hotel to give the maximum score to all items, as suggested by Mellinas et al., (2016) in their research comparing hotels on Booking.com and Priceline, which seems more difficult than getting the maximum score for a single item.

Another reason why hotels with high standards of quality get worse scores on Booking.com than on TripAdvisor could be in how reviews are posted. On Booking.com users post a general review when they decide to enter on the platform, but on TripAdvisor users receive an email asking to rate the experience and to post a review explaining both the pros and cons separately, which obliges users to also think of the negative attributes that might be overlooked when answering in a free format.

The results show that the item with the highest mean is staff, followed by location. Cleanliness, a critical item in customer ratings (Barreda & Bilgihan, 2013) is in third place. Value for money, which generally captures all items to determine the overall

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics of the Booking.com Score</strong></td>
</tr>
<tr>
<td>Booking.com</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

Source: Compiled by the authors based on data from Booking.com.
Table 4
Ratings of Hotels on TripAdvisor

<table>
<thead>
<tr>
<th>Region</th>
<th>Hotels</th>
<th>Mean Score</th>
<th>SD Score</th>
<th>Excellent (5)</th>
<th>Very Good (4.5–4)</th>
<th>Average (3.5–3)</th>
<th>Poor (2.5–2)</th>
<th>Terrible (1.5–1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AME</td>
<td>4,130</td>
<td>3.93</td>
<td>0.57</td>
<td>2.95%</td>
<td>65.11%</td>
<td>28.74%</td>
<td>3.00%</td>
<td>0.19%</td>
</tr>
<tr>
<td>ASP</td>
<td>5,922</td>
<td>3.84</td>
<td>0.58</td>
<td>3.16%</td>
<td>56.70%</td>
<td>36.14%</td>
<td>3.77%</td>
<td>0.24%</td>
</tr>
<tr>
<td>EUR</td>
<td>8,809</td>
<td>3.90</td>
<td>0.62</td>
<td>3.85%</td>
<td>62.52%</td>
<td>28.70%</td>
<td>4.59%</td>
<td>0.35%</td>
</tr>
<tr>
<td>MEA</td>
<td>799</td>
<td>3.91</td>
<td>0.67</td>
<td>4.38%</td>
<td>61.70%</td>
<td>28.29%</td>
<td>5.51%</td>
<td>0.13%</td>
</tr>
<tr>
<td>Total</td>
<td>19,660</td>
<td>3.89</td>
<td>0.60</td>
<td>3.47%</td>
<td>61.28%</td>
<td>30.93%</td>
<td>4.04%</td>
<td>0.27%</td>
</tr>
</tbody>
</table>

Source: Compiled by the authors based on data from TripAdvisor.

Table 5
Ratings of Hotels on Booking.com

<table>
<thead>
<tr>
<th>Region</th>
<th>Hotels</th>
<th>Mean Score</th>
<th>SD Score</th>
<th>Exceptional (10–9.5)</th>
<th>Superb (9.4–9)</th>
<th>Fabulous (8.9–8.6)</th>
<th>Very Good (8.5–8)</th>
<th>Good (7.9–7)</th>
<th>Pleasant (6.9–6)</th>
<th>Review Score &lt;6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AME</td>
<td>4,130</td>
<td>8.11</td>
<td>0.79</td>
<td>1.33%</td>
<td>10.53%</td>
<td>19.30%</td>
<td>32.11%</td>
<td>28.62%</td>
<td>6.59%</td>
<td>1.53</td>
</tr>
<tr>
<td>ASP</td>
<td>5,922</td>
<td>7.67</td>
<td>0.96</td>
<td>0.88%</td>
<td>6.30%</td>
<td>10.74%</td>
<td>24.54%</td>
<td>35.58%</td>
<td>17.44%</td>
<td>4.53</td>
</tr>
<tr>
<td>EUR</td>
<td>8,809</td>
<td>8.07</td>
<td>0.79</td>
<td>1.52%</td>
<td>10.23%</td>
<td>16.35%</td>
<td>33.43%</td>
<td>29.87%</td>
<td>7.07%</td>
<td>1.53</td>
</tr>
<tr>
<td>MEA</td>
<td>799</td>
<td>7.73</td>
<td>1.00</td>
<td>1.63%</td>
<td>7.88%</td>
<td>12.39%</td>
<td>25.41%</td>
<td>32.04%</td>
<td>15.14%</td>
<td>5.51</td>
</tr>
<tr>
<td>Total</td>
<td>19,660</td>
<td>7.95</td>
<td>0.87</td>
<td>1.29%</td>
<td>9.01%</td>
<td>15.12%</td>
<td>30.15%</td>
<td>31.41%</td>
<td>10.42%</td>
<td>2.59</td>
</tr>
</tbody>
</table>

Source: Compiled by the authors based on data from Booking.com.

Table 6
Ratings of Hotels

<table>
<thead>
<tr>
<th>Region</th>
<th>Above 6.25 Booking.com</th>
<th>Below 6.25 Booking.com</th>
<th>3 or above TripAdvisor</th>
<th>Below 3 TripAdvisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AME</td>
<td>97.38%</td>
<td>2.62%</td>
<td>96.80%</td>
<td>3.20%</td>
</tr>
<tr>
<td>ASP</td>
<td>92.20%</td>
<td>7.80%</td>
<td>96.00%</td>
<td>4.00%</td>
</tr>
<tr>
<td>EUR</td>
<td>97.41%</td>
<td>2.59%</td>
<td>95.06%</td>
<td>4.94%</td>
</tr>
<tr>
<td>MEA</td>
<td>90.86%</td>
<td>9.14%</td>
<td>94.37%</td>
<td>5.63%</td>
</tr>
<tr>
<td>Total</td>
<td>95.57%</td>
<td>4.43%</td>
<td>95.68%</td>
<td>4.32%</td>
</tr>
</tbody>
</table>

Source: Compiled by the authors based on data from Booking.com and TripAdvisor.

Table 7
Student’s t Test and Effect Size (Cohen’s $d$)

<table>
<thead>
<tr>
<th>Hotel Category</th>
<th>1 Star</th>
<th>2 Stars</th>
<th>3 Stars</th>
<th>4 Stars</th>
<th>5 Stars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$d$</td>
<td>$t$</td>
<td>$d$</td>
<td>$t$</td>
</tr>
<tr>
<td>AME</td>
<td>-2.52*</td>
<td>-9.63**</td>
<td>-1.56**</td>
<td>0.07</td>
<td>6.31**</td>
</tr>
<tr>
<td>ASP</td>
<td>1.71</td>
<td>5.05**</td>
<td>4.13**</td>
<td>0.08</td>
<td>9.32**</td>
</tr>
<tr>
<td>EUR</td>
<td>-6.18**</td>
<td>-13.54**</td>
<td>-16.27**</td>
<td>0.29</td>
<td>-7.70**</td>
</tr>
<tr>
<td>MEA</td>
<td>0.42</td>
<td>-0.45</td>
<td>-0.34</td>
<td>0.03</td>
<td>5.73**</td>
</tr>
</tbody>
</table>

Note: Coefficients are shown in the table.
*p < 0.05; **p < 0.001.
Fraudulent practices on TripAdvisor cannot be confirmed in this study even though cases are documented in other studies (Mayzlin et al., 2012; Mellinas Cánovas, 2015; Mkono, 2015) because it can be concluded that the verification system on TripAdvisor does not affect the position of hotels when compared with Booking.com because of the strong statistically significant correlation between both rankings for most of the cities.

Thanks to this relationship, hotel managers can roughly know which position their hotels will have on Booking.com based on the position obtained on TripAdvisor, as well as if the scoring scale on Booking.com benefits or harms them. This fact could be interesting information to have before deciding whether or not to work with this online hotel-booking website, as online visibility is important for profitability (Neirotti et al., 2016).

The main contribution of this study is that suspicions of fraud on TripAdvisor because of its unverified user reviews are not the norm on this platform compared with Booking.com, perhaps because the reputation management system on TripAdvisor increases the motivation of users to contribute reliable reviews (Yoo, Sigala, & Gretzel, 2016). As the number of reviews grows, the impact of possible fake reviews falls, as they are overwhelmed by genuine consumer-generated content, thanks to the tendency of human behavior to embrace “the power of the crowd.” Moreover, as stated by O’Connor (2008), TripAdvisor appears to be doing a good job of policing its system.

The distinctive feature of this research is comparing ranking and not scores, since product ranking is now “the strongest antecedent of high-involvement travelers’ adoption of information from online reviews, which is a new finding in eWOM research” (Filieri & McLeay, 2013, p. 52). Moreover, it is important to note that this study extends the number of investigations analyzing a very large sample of hotels in more than 400 cities. As Xiang, Schwartz, Gerdes, and Uysal (2015) pointed out, there are very few studies in this field that explore the capacity of a large amount of data.

A lack of veracity is associated with TripAdvisor in some studies (Gerrard, 2012; Morrison, 2012; Fuchs & Zanker, 2012), is the second most underrated item in this study. The highest average of all items is location. An explanation for this fact might be that the traveler already knows where a hotel is located and therefore it is the item that generates less uncertainty on the trip and less surprise in the experience. It is worth noting that the geographical location of a hotel has the highest average of all items, essentially because it cannot be changed or improved by the staff once the property is built (Xie et al., 2014). Of the 20,000 hotels, none had a score below 3.3 for location and none below 3.6 for staff.

With the overall rating of hotels on Booking.com, no hotel had a score less than 3.7; it did not happen on TripAdvisor, where some hotels had the minimum score. Only 4.43% of the hotels had a score below 6.25 on Booking.com, which, on a scale from 2.5 to 10, is in the middle. Because the users of this system are unaware of this fact, it may favor their perception that most hotels on Booking.com are above 5, a score that is in the middle of a typical scale from 0 to 10. Mellinas et al. (2015) pointed out that, “this scoring system does not appear to be illegal but could indicate a lack of honesty with customers” (p. 74).

TripAdvisor achieves a score below 3 (midpoint on a scale from 1 to 5) in 4.32% of the cases, so it seems that the number of hotels rated poorly on both websites is similar. Although Bjørkelund et al. (2012) noted that scores from both sources shift toward the higher end of the scale, Booking.com scores are noticeably higher, with more than 80% of review scores higher than 6 on a scale from 0 to 10. Bjørkelund et al. (2012) did not detect that the scale is from 2.5 to 10 and other studies conclude that there is a positive bias on Booking.com when compared to TripAdvisor. This bias could be due to the fact that TripAdvisor offers “a valuation related more to ‘real’ quality (services offered) than to just administrative category or price” (Fernández-Barcala et al., 2010, p. 358). However, the results show a positive bias on both Booking.com and TripAdvisor, and the review scores on both websites are found to be greatly skewed from the middle to the top of the scale, in line with a study conducted on FlipKey.com reviews (Racherla, Connolly, & Christodoulidou, 2013).
would be optimal to ensure that opinions more accurately reflect the reality of each hotel. Customer opinions are also a source of segmentation that allows the better positioning of each hotel (Martin-Fuentes, Fernandez, & Mateu, 2016). It is not beneficial to cheat users with fictitious reviews about the hotel because future guests will be disappointed if the hotel does not meet their expectations.

This study shows that when opinions are given on different websites with different rating and user verification systems, the outcome in terms of ranking ultimately tends to be the same, although there are differences in how to collect the reviews. Booking.com sends an email asking for a review within the following 28 days after the stay, and TripAdvisor users can opine whenever they want. The rankings are similar on each site also despite the differences in the scoring scales and in the survey methods (free format or pros and cons separately), and the antiquity of the ratings (TripAdvisor calculates the ratings based on all reviews received while Booking.com does not take into account the reviews that are more than 24 months old). Our research is a grounded work that allows us and other researchers to deepen the subject on suspicious unverified users on TripAdvisor and on the measuring scales in sales and advice platforms.

Although hotels all over the world were analyzed in this study, using data from Booking.com and TripAdvisor, these websites can produce a cultural bias because they are used by some nationalities more than others. Empirical replications using other channels (e.g., Ctrip) to determine if there are behavioral differences by nationality may provide more insight to this discussion.

Acknowledgments

This work was partially funded by the Spanish Ministry of the Economy and Competitiveness: research projects ECO2017-88984-R, TIN2015-71799-C2-2-P, and ENE2015-64117-C5-1-R. This research article received a grant for its linguistic revision from the Language Institute of the University of Lleida (2016 call).

References


Webb, S. (2014). ‘Give us free meals or I’ll put a bad review on TripAdvisor’: Anger at blackmailers using online ratings website to extort free rooms and dinners


