Average Scores Integration in Official Star Rating Scheme

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Journal of Hospitality and Tourism Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>JHTT-07-2017-0050.R3</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Refereed Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>eWOM, Hotels, Average Scores, Star Rating Systems, Online Reviews</td>
</tr>
</tbody>
</table>
When revising your paper, please prepare this report explaining how you have responded to each reviewer’s comments and suggestions specifically.

<table>
<thead>
<tr>
<th>Suggestions/comments from the Reviewer</th>
<th>Response from the Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Make sure to follow JHTT author guidelines closely: <a href="http://www.emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=jhtt">http://www.emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=jhtt</a> For example, when there are three or more authors, you need to use Adam et al., XXXX (or Adam et al., XXXX) format for the first time and after.</td>
<td>The paper has been duly reviewed to ensure that the JHTT author guidelines have been met. Particular attention has been made to the in-text citation format for when three or more authors.</td>
</tr>
<tr>
<td>2. Revisit the Introduction to highlight the research purpose, research problem, and significance of the study. Please explain these in application to both hotels and hotel guests. Revisit the Discussion and Conclusions sections one more time to better answer the &quot;So What&quot; question. There should be four sub-sections under this section: (1) Conclusions, (2) Theoretical Implications, (3) Practical Implications and (4) Limitations and Future Research.</td>
<td>Explicit mention to the research purpose, research problem and significance (i.e. implications of the study) for both hotels and guests have been included in the introduction. We have revisited the Discussion and Conclusion sections to better answer the “So What” question and have divided the section in the requested 4 subsections.</td>
</tr>
<tr>
<td>3. Cross check all references within text with your reference list. You may like to add more recent and relevant references published in recent months/years in JHTT or other journals.</td>
<td>A thorough check that all the in-text citations are in the reference list has been made, and additional references from work published in recent months/years in relevant journals (including JHTT) have been added.</td>
</tr>
<tr>
<td>4. Keep your article between 6000 and 7000 words including references, tables and figures.</td>
<td>The article is 6501 words including abstract, figures and references.</td>
</tr>
<tr>
<td>5. Proofread your article one more time and also you may ask a technical writer/copy editor to proofread it for you. After the manuscript is accepted, we will not ask you to proofread it again. In short, after I send you an official acceptance e-mail, you will not be able to make any further changes in your manuscript.</td>
<td>We have proofread the article one more time and asked a technical colleague to proofread it for us.</td>
</tr>
<tr>
<td>6. Submit a clean version of your paper. You don’t need to show/highlight all the changes made in the paper. I will read its final version anyway.</td>
<td>OK, thank you. We are submitting a clean version of the paper as suggested.</td>
</tr>
<tr>
<td>7. Include a brief report showing how you have responded to the above requests. You don’t need to show/highlight all the changes made in the paper. I will read its final version anyway.</td>
<td>We are including this report in response to this request.</td>
</tr>
<tr>
<td>8. Finally, take the above suggestions seriously when revising the article. Failure to do so may result in an unfavorable decision as we do not really have time to work on an article multiple times.</td>
<td>We have taken all these comments very serious and made all the amendments to the best of our abilities. We hope that this version of the paper fulfils all the requirements for its publication in the JHTT.</td>
</tr>
</tbody>
</table>
Average Scores Integration in Official Star Rating Scheme

Abstract

Purpose of this paper: Evidence suggests that Electronic Word of Mouth (eWOM) plays a highly influential role in decision-making when booking hotel rooms. The number of online sources where consumers can obtain information on hotel ratings provided has grown exponentially. Hence, a number of companies have developed Average Scores to summarise this information and to make it more easily available to consumers. Furthermore, Official Star Rating schemes are starting to provide these commercially developed Average Scores to complement the information their schemes offer. This raises questions regarding the robustness of these systems, which this paper addresses.

Design/methodology/approach: Average Scores from different systems, as well as the scores provided by one rating site were collected for 200 hotels and compared.

Findings: Findings suggested important differences in the ratings and assigned descriptive word across websites.

Research limitations/implications (if applicable): The results imply that the application of Average Scores by Official organisations is not legitimate and identifies a research gap in the area of consumer and star rating standardization.

What is original/value of paper: The paper is of value to the industry and academia related to the examination of rating scales adopted by major online review tourism providers. Evidence of malpractice has been identified and the adoption of this type of scales by Official Star Rating schemes is questioned.

Keywords: eWOM, Hotels, Average Scores, Star Rating Systems, Online Reviews

1. INTRODUCTION

Social media in general, and Electronic Word of Mouth (eWOM) in particular, have completely restructured business consumer relationships. eWOM refers to non-formal online communication about products, services or their sellers (Litvin et al., 2008). A growing number of websites allow consumers to post their opinions about products, contributing to eWOM. This information can then subsequently be accessed by other buyers and considered at the time of making purchases.

Due to the growing number of eWOM sites, and overwhelming amount of information, systems synthesizing this information for consumers have emerged. These systems produce average scores from the different eWOM sites.

This information is highly valuable, and this is the reason why a number of official bodies, namely hotel star accreditation systems are starting to integrate these average scores into their own accreditation systems (Blomberg-Nyard and Anderson, 2016). However, the problem is that the process by which these average scores are produced lack transparency, which is one of the fundamental characteristics of star accreditation systems.

This raises questions about the robustness of these scores. Therefore, the purpose of this study is that one of examining the robustness of these scores. To do this, three key questions will be addressed: i) whether the different average scores provided by license software are resulting in equivalent or different scores. In the case of a difference, hoteliers with strong lobby capacity could support the adoption of that system which provides them with a more favorable score. And an example of this can already be found in the work by UNWTO (2014), which explains that the system developed for Norway has not been applied due to opposition
from major hotel chains. ii) whether the "average score" provided by these software systems is different to those provided by the metasearch websites. In the event that the differences are minimal, the adoption of licensed pieces of software, which are likely to be quite costly, would not be justifiable. Finally, iii) if significant differences among systems were observed, it would be of interest to determine the reasons.

The results will have strong implications for hotels and hotel guests. Findings suggesting lack of robustness would justify the position of those hotels challenging the integration of these average scores from online reviews into official star classification accreditations. For hotel guests it would imply that consumer information and possibly new policies would be required to ensure consumer protection.

2. LITERATURE REVIEW

eWOM is more influential than traditional WOM due to its speed, convenience, ability to reach many and the lack of human pressure which influences face to face communication (Sun et al., 2006). Additional reasons for consumer attention to eWOM relate to the expectation that receiving information may decrease the time and effort taken in making a decision and/or contribute to the outcome of making a more satisfying decision (Schiffman and Kanuk, 2000).

This breadth of eWOM scope and ease access can deeply affect a company’s performance. Companies who adequately manage eWOM can have a competitive advantage (Loureiro and Kastenholz, 2011). This is the reason why companies are increasingly seeking to understand the factors that influence the use of eWOM and its impact.

The influence of eWOM on tourism has been studied extensively. Femback and Thomson (1995), Wang et al., (2002), and Yoo et al., (2009) suggest that travel online reviews are perceived as similar to recommendations by friends and relatives, and as more trusted information than the official one. Gretzel et al., (2006), and Wang et al., (2002) suggest that the reason for this impact is that social media decreases uncertainty, and it provides a sense of belonging into virtual travel communities. Other studies argue that the influence of eWOM on purchase decisions, may also be related to how the ratings position them on the ranking of searching results (Reino and Massaro, 2016). Thus, those establishments with the highest scores appear first in the searching results.

In the hotel sector, online reviews and ratings have a significant impact on potential consumers and their purchase decisions, both in terms of number of bookings (Onur Taş, 2012), possibility to increase room price (Ye et al., 2009; and Anderson, 2012) and increment of occupancy rate (Vigilia et al., 2016).

2.1 Average Score

An increasingly growing number of Internet sites provide access to hotel scores. The average score facilitates the access to this information from across sites, synthesizing it into a single web. Furthermore, it enables drawing comparisons between a larger number of hotels and the analysis of their evolution over time.
Currently, there are mainly two business models behind the implementation of average score systems for commercial purposes. The first one refers to commercialisation tools that help hoteliers understanding their online reputation. This is the case of ReviewPro (www.reviewpro.com), TrustYou (www.trustyou.com) and Olery (www.olery.com). These software vendors perform a compilation of reviews from dozens of sources of information on hotels, offering detailed reports on valuations of hotels by department, by keyword, client profile, etc. It is specialized software, which has an economic cost for businesses, but can provide valuable information. The services offered by these companies generally relate to 1) pulling together reviews and ratings in one dashboard from different review platforms; 2) integrating and weighting scores through an algorithm based on a number of variables and providing a holistic score, typically on a scale from 1-100; 3) comparing hotel performance within a group, or a competitive set; 4) live monitoring whereby push notifications can be set for ratings below a predefined threshold; and 5) the provision for hoteliers to respond to reviews in one platform. However, a license is required to use this software (Henses, 2015).

The level of information provided prior payment differs across systems. For example, TrustYou only facilitates certain information for free (i.e. the global average score) but most features are restricted and only available after payment. However, Olery does not provide access to any information without having taken on the software license previously. ReviewPro did not use to allow free access to their scores. However, from 2015 onwards, the company started providing free access through some official organisations. Further details can be found in the following section.

The second type of systems relate to metasearch sites. These provide a comparison of hotels with a summary of their inventory available through different online booking platforms. Examples include Kayak (www.kayak.com), Trivago (www.trivago.com), Skyscanner (www.skyscanner.com) or HotelsCombined (www.hotelscombined.com). Along with price and other information, metasearch websites show an average valuation of the property based on the reviews of guests on other sites. The information provided by metasearch websites is freely available and its data has been referred to in a number of research papers in the field. This is the case of Schamel, 2012; Pouplana, 2014; and Pesonen and Palo-oja, 2010.

2.2 Official Use of the Average Score

UNWTO (2014) and Blomberg-Nyard and Anderson (2016) suggested the need to integrate online review ratings into hotel classification. This would be done to complement the quantitative measures offered by hotel classification schemes with the qualitative information provided by online reviews. In their investigation of online reviews and star ratings, these authors found no correlation, and that they serve complementary purposes. Furthermore, these authors refer to two industry studies (one by the National Tourism Development Authority of Ireland and another one by Tourism Ireland), which concluded that online ratings are considered more important than star ratings by both consumers and hoteliers.

A number of official organisations (including tourism boards, accrediting bodies and trading organisations) have started providing access to Average Scores. They provide it in collaboration with commercialisation tools such as those outlined in the previous section. An example is the Australian Star Rating scheme. They facilitate this information through their own website.
and use ReviewPro average rating scores about the hotels listed to complement the star rating information. The Switch Hotel Association is also providing similar type of data through their website (https://hotels.swisshoteldata.ch/) and have developed this through a collaboration with TrustYou. Though according to Thiessen (2013) and Edgcumbe (2014), Abu Dhabi was the first destination to incorporate online ratings and feedback into its classification system. This was done in collaboration with Olery, a tourism data management provider. An additional destination providing this type of information is Bahrein (Hensens, 2015). And Norway, with its model developed by QualityMark Norway, and Germany are working towards the development of similar data (UNWTO, 2014), though limited details are available.

The approach of resorting this type of service to an existing software company seems understandable. After all, it involves a high volume of data and it would be an overwhelming task for Official Organisations if wanting to do it by themselves. The problem is that these specialised software companies fail to provide an explanation of the methodology used. When requested to the software companies, this is referred to as a "secret algorithm" or a "proprietary algorithm". However, when these scores are adopted by Official Bodies they acquire an “official” character, and it would be reasonable to expect that the algorithms become publicly available, providing transparency in their rating process.

Academic literature has previously suggested the importance of standardizing this type of data prior to its aggregation (i.e. Chaves et al., 2012; Tano et al., 2014; and Viglia et al., 2016). Nevertheless, even these studies fail to reach a consensus on the establishment of an agreed methodology.

3 METHODOLOGY

Secondary data (quantitative and qualitative), has been used to address the three research questions of this study, which are:

1) Are there variations in the average scores provided by the different licensed software companies?
2) Are there differences between the average score provided by software systems and metasearch websites?
3) In the event that significant differences were observed, what are the reasons?

Quantitative Data was collected and analysed in order to answer questions 1 and 2. The most reasonable approach would be to take a random sample of hotels and to check the scores obtained by them in each of the different systems (including both licensed software companies and metasearch websites). However, as noted in the previous section, of the 3 licensed software companies, only TrustYou provides free access to average scores for all their listed establishments. ReviewPro scores for Australia are freely available through the Australian Rating Scheme (https://www.starratings.com.au/) and access to ReviewPro data for Spain was available through Travel Advisors (www.traveladvisorsguild.com). Therefore, the sample was developed taking into account these limitations. Data was extracted from the two licensed software companies for which data was available (i.e. TrustYou and ReviewPro) and four hotel metasearch (HotelsCombined, Trivago, Skyscanner and Kayak) for Australian and Spanish hotels only. In addition, data from Booking.com (one major online booking systems that
provides verified opinions of hotels) was also collected for the same establishments, to reveal any possible differences between aggregating systems and single scoring sites.

A random stratified sampling approach was adopted. The search took place in an aleatory manner but fulfilling the following two conditions: they all have information emerging from at least 7 web sources and they all have at least 200 opinions in each of the webs. This criterion was established to ensure that the comparison of data across sites would be feasible and consistent. Searches through the database were undertaken in the major cities of each country (19 in total) in January 2016. One hundred Spanish hotels and the same number of Australian ones listed in ReviewPro were selected.

Data needed to be standardised. All the systems under study have one of the following types of scales: a scale with 100 as the maximum score and whole numbers or a scale with 10 as the maximum score and decimal values. Both scales are comparable when multiplying by 10 the values of the second scale. Therefore, this approach was taken.

It was decided that in the event that answering question 3 was applicable, this would be done by undertaking an analysis of the descriptions provided by the different systems in their own websites and the characteristics of their reviewing scores systems. However, given that these systems provide limited information about the way their means are calculated, it may not be possible to undertake a systematic evaluation of the same systems included in the previous analysis. Therefore, a heuristic approach has been adopted, in which different systems, including those in the previous part of the analysis, but also others, have been included. This relates to the two licensed software companies for which data was available (i.e. TrustYou and ReviewPro); and the four hotel metasearch sites (i.e. HotelsCombined, Trivago, Skyscanner, Kayak); as well as Booking.com. Additionally, data from Olery, which is the aggregator providing the ratings for Abu Dabi, as well as Agoda and TravelRepublic (both OTAs), which are included in the two UNWTO documents advocating for the integration of these systems in official star rating schemes (UNWTO, 2014), have also be included in the analysis. Finally, as previously explained, Olery was not included in the previous parts of the analysis due to the lack of public accessibility to their average scores.

4 RESULTS

Data from 200 hotels from the 7 different systems was collected, producing a total of 1,400 average rating scores. The typical number of reviews from which each of these average scores emerged was 397,124. However, the specific numbers varied across systems. For example, it was noted that Trivago’s average scores were based upon over three times the number of reviews used to produce ReviewPro’s average scores (666,984 reviews in Trivago vs. 212,972 in ReviewPro).

The results show different average scores across systems. Only one of the establishments in the sample of 200 hotels received the same scores across systems. Furthermore, while in 20 cases the differences were just one point, in the remaining cases (180), these are two or more
In addition, there are 67 cases (33.5%) in which the difference is equal to or greater than 5 points and two of the cases show the very substantial difference of 9 points.

4.1 Differences Among Licensed Software Providers

Table 1 shows the overall comparison of average scores received by the hotels in the sample. This data shows that the difference between ReviewPro and TrustYou is limited (lower than 1). The t-test analysis corroborates that the difference is small by suggesting that it is not significant (p>0.05). However, it should be noted that the scores differ for the large majority of the establishments (73%). The difference is determined by 62% of the establishments for which the scores are higher in ReviewPro, and 11%, for which the scores are higher in TrustYou. Only 54 of the hotels (27%) display exactly the same scores.

Additionally, these differences significantly affect the rankings of hotels, as shown in table 2. When reordering the ranking of 100 Australian hotels with the TrustYou scores, only 15 occupy the same position, while 45 worsen it and 40 improve it. Of those improving their position, there are 14 that improved it in 5 positions or more, and 15 which worsened it in 5 positions or more. Similarly, when the ranking of Spanish hotels is reordered with the scores given by TrustYou, only 13 establishments stay in the same position, 45 improve it and 42 worsen it, with 25 hotels seeing their position changed by 5 or more positions.

Therefore, in terms of answering the first question of this study, i.e. “are there differences in the average scores provided by the different licensed software companies?”, the answer is yes, there are differences in the average scores provided by licensed software companies. And these differences do not only relate to the average scores obtained by establishments throughout the different systems, but also to their position in the ranking of results. This has important implications for the hotel industry, as it may imply that powerful corporations will lobby for their countries’ star classification systems to adopt those systems which benefit their position in the rankings.

As explained in the literature review section, not only are scores important in determining consumer’s choice (Öğüt and Onur Taş, 2012; Ye et al., 2009; Anderson, 2012; Viglia et al, 2016) but also the position in the rankings in which hotels appear when users undertake a search (Reino and Massaro, 2016). Therefore, based on the differences presented across system, there is a strong reason to believe that the performance of hotels will vary depending on which licensed software is adopted. This suggest that despite requiring payment, integrating licensed software providers, instead of other type of free aggregators (e.g. metasearch sites), offers no guaranteed of the reliability of the average scores that they
provide. And given the lack of transparency of both these systems, it is not possible to find out which one of the two providers is reliable, if any.

4.2 Differences Among License Software Systems and Metasearch Websites

Looking back at table 1, this shows that there are not important differences in the average scores produced across systems, except Trivago’s. Trivago’s results are lower than all the others. As an illustration of this, the difference between Trivago (which provides the lowest score) and ReviewPro (which provides the highest) is of 2.01 points.

TABLE 3 GOES HERE

Independent T-test was carried out on all paired systems (see table 3), and the results suggested significant differences (p<0.05) across a considerable number of paired systems.

In terms of differences between the licensed systems and the metasearch sites, these exist. Review-pro presents significant differences when compared with Trivago (p<0.001), Skyscanner (p<0.05), Kayak (p<0.05). Trustyou is the one presenting differences with the least number of other systems, and this is only with trivago (p<0.05). And this gives reasons to believe that of the two licensed systems, this may be the most reliable one.

With regards to the reliability of the metasearch systems, Trivago shows significant differences with the two licensed systems (ReviewPro and TrustYou, p<0.001 and p<0.05 respectively), and also with one of the metasearch systems, i.e. HotelsCombined (p<0.01). However, HotelsCombined only presents differences with Trivago (p<0.01). Similarly, Skyscanner and Kayak, only present differences with one other systems. Interestingly, they both present differences with ReviewPro (both p<0.05).

Based on these analysis, there is a strong reason to believe that the average scores provided by HotelsCombined, Skyscanner and Kayak (all freely available), may offer similar reliability, at least, to those provided by TrustYou, and more trustworthiness than those provided by ReviewPro. This has important implications in terms of their possible integration by official rating schemes, which relates to the fact that both TrustYou and ReviewPro are licensed systems, and their use requires payment; while the metasearch sites are freely available.

Consequently, in answer to the second question of this study: yes, there are significant differences between the average score provided by these software systems and those by metasearch websites. However, further analysis suggested that differences may not necessarily be related to the lower reliability of metasearch websites. This has important implications for hoteliers and also for star rating classification systems, which may be considering the adoption of average scores. Taking this into consideration, and the fact that metasearch websites are free and licensed systems require a subscription, the adoption of the latter may be difficult to justify.

4.3 Possible Influencing Factors
An analysis of the descriptions offered through the different systems' websites has suggested possible factors generating these differences. Information looked for included sources of data; any possible difference related to data which is excluded (e.g. old reviews); and differences in the way the data may be transformed to new scales.

As explained in the methodology section, this analysis included the same systems which were examined in the previous two subsections (i.e. TrustYou, ReviewPro, HotelsCombined, Trivago, Skyscanner, Kayak and Booking.com), and this was completed with others such as Olery, Agoda and TravelRepublic. However, no description of the way this data is sourced and/or produced has been found for HotelsCombined, Trivago, Skyscanner or Kayak. Some details about ReviewPro were found through the StarRating Australia website, as well as some others about TrustYou, and these form part of the analysis below. Additionally, data about Olery, Agoda and TravelRepublic has also been included to support the analysis and discussion.

Through the analysis, the first difference that has been noted, relates to the time periods for which online reviews and ratings are kept. During the examination of the sites, it was noted that Booking.com deletes all the reviews which are older than 24 months, while other sites don’t (e.g. TripAdvisor which is one of the sources of Olery). Olery explains on their site (Olery, n.d.) that “only the reviews from the last three months are taken into account, and newer reviews are weighted a bit more heavily than older reviews”. However, no explanation of how these weights are generated is provided.

The second difference identified relates to the frequency of updates. As described by StarRating Australia on their site (Star Rating, n.d.) “…the Travellers’ Rating on starratings.com.au [which is provided by Reviewpro.com] will be updated monthly…” however, all the other systems fail to provide any information regarding the frequency with which their ratings are updated. This suggest differences in the systems and websites’ policies about keeping old reviews. And this is likely to have an important effect on the results.

The third difference identified relates to the use of different scales. As explained by UNWTO (2014), the variety of websites that collect scores use different scales, and these must be transformed to a single scale. For example, Booking.com uses a 2.5-10 scale (Mellinas et al., 2016), Agoda uses a 2-10 and TravelRepublic uses a 0-10. And interestingly, UNWTO (2014) assumes that Booking.com and Agoda use a 1-10 scale; error that could lead to inaccuracies in final average score or divergences between different score aggregators. The process by which each average score aggregator undertakes this transformation could well determine that difference in the results. ReviewPro conducted and published an analysis of the scales used by the different sites where the reviews are originated. However, this is not the case of all the other systems, which have not provided any type of explanation to the process by which ratings are transformed. And even ReviewPro fails to make full disclosure.

The fourth difference identified relates to the sources from which comments and scores are extracted. For example, ReviewPro suggests that they take into consideration social sites like Facebook while that is not the case of all the others. On the other hand, Olery for example takes into consideration the unverified ratings from TripAdvisor, while others such as ReviewPro and TrustYou do not. Given that TripAdvisor holds over 300 million reviews, their addition/inclusion to the data is expected to have an important impact on the results. The case
of Olery is interesting, as they suggest that “...ratings from Booking.com are weighted more heavily in the GEI calculation than those from TripAdvisor...” (Olery, n.d.), due to the fact that ratings from Booking.com are verified. This suggests further problems for data standardisation across systems though.

These differences have important implications for consumers, as it suggests that the information that they receive is not robust and may be calling for the need to develop policies to ensure consumer protection.

5 CONCLUSIONS

The tourism industry is entering a process of introducing online hotel ratings official as part of the standard hotel classification, either in an integrated or complementary form. The aforementioned report UNWTO (2014), and the first experiences in Australia, Abu Dhabi, Switzerland, Germany, Bahrain and Norway show that this practice has already started to be deployed at a practical level. The intention of this type of initiatives is that one of providing an official status to this type of score. This type of status will automatically grant it a higher level of recognition than those rankings currently provided directly by private entities such as TripAdvisor, Expedia or Booking.com. It is also explained in the article that slight variations on online ratings of hotels significantly influence their occupancy levels. Therefore, this article has aimed to find out whether the data provided by the different average score aggregators is robust enough to make this type of adoption legitimate.

The results have shown that there are important differences not only in the rating scores obtained by the hotels but also on the rankings in which the hotels appear when undertaking a search. And this may have a more important effect than the rating itself, as it acts as a filter when searching for hotels.

The analysis has demonstrated that there are variations on the average scores provided by the different licensed software providers. Furthermore, they have also manifested differences between the group of licensed software providers and that one made up by metasearch websites. Further analysis of the systems has concluded that variations may be determined by the differences in the time periods during which the data is collected; variations in the frequency of data capture; inconsistency in the type of scale adopted and the process by which these are transformed; and differences in the sites which each aggregator includes. Therefore, it seems that the data provided by average score aggregators (either licensed or free metasearch sites) is not robust enough for their adoption by official organisms, as they stand at present.

5.1 Theoretical Implications

This study builds from extant knowledge on eWOM impact and use by hotel organisations, as well as by official star rating schemes for hotels. The findings expanded on this knowledge by analysing and explaining their practices.
5.2 Practical Implications

The findings about the differences in the rankings has important implications for the hotel industry, as it may imply that powerful corporations will lobby for their countries’ star classification systems to adopt those systems which benefit their position in the rankings.

The differences in the ratings have important implications for consumers, as it suggests that the information that they receive is not robust and may be calling for the need to develop policies to ensure consumer protection.

Additionally, findings in the lack of robustness of these average scores also have important implications for star rating classification systems, which may be considering the adoption of average scores. This study opposes the calls made by UNWTO (2014) to support the integration of these systems in official rating scales until further transparency is provided by these systems. Furthermore, in the event of ignoring this recommendation and deciding to integrate average scores as they currently operate, it is herewith argued that the practice of integrating a paid system such as ReviewPro or TrustYou does not present sufficient justification. The comparison across systems provided through this study has given no indication to believe that the average scores produced by these two licensed systems offers additional reliability than the average scores provided by metasearch websites such as HotelsCombined, Kayak and Skyscanner, which are all freely available.

5.3 Limitations

The limitations of this study mainly relate to data access. For the last part of the analysis, information about certain systems was not available. Therefore, further sites needed to be integrated in the analysis to supplement this.

5.4 Future Research

Despite the issues of lack of transparency and robustness presented by these systems, there are already a few cases of their integration in star rating systems by official organisations. If so far this issue has not generated any controversy it could be due to the low implementation of the system and the relative lack of attention paid to it by consumers. At the time when this practise becomes generalised, surely hoteliers will be interested in finding out how the score is assigned to their establishment. The only way to avoid problems is to establish a transparent system for calculating overall scores, with an algorithm of public and open character. It would also be advisable that the criteria used were previously agreed with hoteliers, thus facilitating its subsequent acceptance by the sector.

Therefore, future research should be undertaken in the development of a robust methodology to ensure the robustness of the aggregation techniques used by these sites. In this process, it may still be advisable to count with the collaboration of software suppliers (Olery, ReviewPro, TrustYou), since the amount of information that needs collecting and the underlying processes may require using their technology. It would also be desirable to try avoiding the disparity of criteria by countries currently exists with regard to systems of hotel classification (stars,
diamonds, etc ...), so that the initiatives led by UNWTO, EU or other transnational organisms, are more suitable.

Finally, consumer perceptions on the practice of aggregating this data to official star rating classifications would also be very valuable.

REFERENCES


Table 1. Average Scores Received by the Hotels

<table>
<thead>
<tr>
<th>TYPE OF SYSTEM</th>
<th>LICENSED SYSTEMS</th>
<th>METASEARCH SITES</th>
<th>INDIVIDUAL SCORE PROVIDER</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WEB</td>
<td>REV.PRO</td>
<td>TRUSTYOU</td>
<td>H.COMB</td>
</tr>
<tr>
<td>SCORE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>82.79</td>
<td>82.03</td>
<td>82.24</td>
<td>80.78</td>
</tr>
<tr>
<td>NUMBER OF REVIEWS</td>
<td>1064.86</td>
<td>2265.44</td>
<td>1883.60</td>
<td>3334.92</td>
</tr>
</tbody>
</table>
Table 2. Position in the TrustYou Ranking Compared to ReviewPro

<table>
<thead>
<tr>
<th></th>
<th>EQUAL</th>
<th>IMPROVE</th>
<th>WORSEN</th>
<th>IMPROVED 5 OR MORE POSITIONS</th>
<th>WORSEN 5 OR MORE POSITIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUSTRALIA</td>
<td>15</td>
<td>45</td>
<td>40</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>SPAIN</td>
<td>13</td>
<td>45</td>
<td>42</td>
<td>13</td>
<td>12</td>
</tr>
</tbody>
</table>
### Table 3. Independent T-Test

<table>
<thead>
<tr>
<th></th>
<th>REV.PRO</th>
<th>TRUSTYOU</th>
<th>H.COMB</th>
<th>TRIVAGO</th>
<th>SKY</th>
<th>KAYAK</th>
</tr>
</thead>
<tbody>
<tr>
<td>REV.PRO</td>
<td>0.138155262</td>
<td>0.3203985</td>
<td>0.0001</td>
<td>0.04031</td>
<td>0.0275</td>
<td></td>
</tr>
<tr>
<td>TRUSTYOU</td>
<td>0.138155262</td>
<td>0.7127957</td>
<td>0.01464</td>
<td>0.56911</td>
<td>0.46395</td>
<td></td>
</tr>
<tr>
<td>H.COMB</td>
<td>0.320398513</td>
<td>0.712795673</td>
<td>0.00924</td>
<td>0.37337</td>
<td>0.29911</td>
<td></td>
</tr>
<tr>
<td>TRIVAGO</td>
<td>0.000101917</td>
<td>0.014637139</td>
<td>0.0092429</td>
<td>0.05953</td>
<td>0.08723</td>
<td></td>
</tr>
<tr>
<td>SKY</td>
<td>0.040311756</td>
<td>0.569113884</td>
<td>0.3733748</td>
<td>0.05953</td>
<td>0.86782</td>
<td>0.86782</td>
</tr>
<tr>
<td>KAYAK</td>
<td>0.027502688</td>
<td>0.463949955</td>
<td>0.2991134</td>
<td>0.08723</td>
<td>0.86782</td>
<td>0.86782</td>
</tr>
<tr>
<td>BOOKING</td>
<td>0.116537488</td>
<td>0.975653735</td>
<td>0.6835171</td>
<td>0.01229</td>
<td>0.57495</td>
<td>0.46543</td>
</tr>
</tbody>
</table>