

Reviewer #1 Response

Thank you for your review observations. I will quote each observation and respond to it while indicating my corrections in the text. Aside from these corrections, we also performed a general grammar check and proofread to the new manuscript, including unchanged sections. Because of comments by reviewers, we redid the experiment with more balanced data, so the results shown are different in their quantitative values. However, our conclusions and hypothesis testing remained largely unchanged by this. We also added a new research rationale and focus, which meant new hypotheses were added to the manuscript. The new manuscript focuses on how these populations interact with the high-standard spirit of Japanese hospitality and service called Omotenashi. In this environment, will the cultural differences be strong enough that their satisfaction stems from different factors?, or will both populations be largely satisfied by service and staff related terms?

1. Machine learning

While this manuscript contains machine learning in its title, there is no major discussion about it. What is machine learning? How do the authors define it? Is data mining part of machine learning? If so, how so? Why is important or useful to use machine learning particularly for this study? What do the authors want to achieve with machine learning?

Machine learning and data mining often overlap, which is why we had an oversight in defining both of them and explaining their differences. In a journal about Information Technology, we assumed it was general knowledge, but we added it for consistency and for a more widespread reader understanding. In the section previously titled 3.4 "Data mining, knowledge discovery and sentiment analysis", thus renamed 3.4 "Data mining, machine learning, knowledge discovery and sentiment analysis" we introduced two paragraphs defining machine learning, its importance of use for analyzing larger samples, and what we can achieve with it. However the manuscript title changed from another reviewers request.

Also, the manuscript can be improved by justifying and elaborating on the necessity of a large sample. Why is it so important about analysing a large sample? What problem did a "small" sample used in existing literature cause? I recommend the authors to prepare a section dedicating to these questions.

In regards to the sample size, a paragraph in previously 3.2, now section 3.3 "Chinese and Western tourist behavior" already describes the importance of these bigger samples. We added a bit of text to expand on this importance.

"Although the studies focusing only on Chinese tourists or only on Western tourists have a narrow view, their theoretical contributions are valuable. We

can see that depending on the study and the design of questionnaires, as well as the destinations, the results can vary greatly. Not only that, but while there seems to be some overlap in most studies, some factors are completely ignored in one study but not in the other. However, since our study uses data mining, the definition of each subject or factor is left for the hotel customers to decide en masse via their reviews (instead of being defined by the questionnaire). This means that the factors will be selected through statistical methods alone. This opened opportunities to find factors that the writers of this study would not have contemplated, or avoid enforcing a factor on the mind of reviewers by presenting them with a question that they didn't think of by themselves. This large variety of opinions in a big sample, added to the automatic findings of statistical text analysis methods, is what gives an advantage to our study in comparison to others with smaller samples. In addition, this study could help us analyze the satisfaction and dissatisfaction factors cross-culturally and compare them with the existing literature."

2. Theoretical discussion on culture I also recommend the authors to develop the story on culture with theoretical perspectives. Why would Westerners show different preferences from Chinese? Why do Westerners value soft attributes (or managerial factors) like staff more while Chinese value hard attributes (or environmental factors) like size? Why do both Westerners and Chinese share dissatisfactory attributes while they show different satisfactory attributes (see Table 6)? While this study did not measure "culture" per se, I believe the authors can find theoretical explanations.

Another reviewer pointed out this, as well as suggesting a new research rationale. Because of it, we added new content to the introduction, changed the abstract and titles, as well as adding a new section in the discussion with theoretical perspectives on the context of Japanese hospitality. It is right that our study did not measure "culture", so we can't be sure of the explanations to certain aspects, although we do explain in the literature review that it has been previously studied how different backgrounds and cultures result in different expectations and as such, different satisfaction factors.

In summary, our expanded discussion is about the differences between Chinese and Western tourists in the context of Japanese hospitality. Western tourists had hospitality and service related satisfaction keywords at the top of their priorities. However, Chinese tourists did not. We theorize this can be from two different causes. One theory is that the Chinese customers have different expectations based on their culture background, and that these expectations lead to different reactions. Specifically, not expecting something and then getting it could be less satisfying than getting an expectation met. Another theory is that despite being in the environment of Omotenashi, their language barriers are enough to disregard hospitality because of their unease or disappointment with the staff and communication availability.

We also expanded on the discussion section by adding more to the previously 7.5, now section 7.2 Managerial vs. Environmental satisfaction, as well as add reasoning to each discussion section as well. In this section we added potential ramifications of the distribution of these factors. While Westerners can be satisfied even more by improving on managerial factors, a big portion of Chinese satisfaction comes from environmental factors. This means that it could be harder for managers to achieve an improvement in their satisfaction. However, we also claim that improving on cleanliness and making hotels tobacco-free could be a solution that benefits both populations. If this is done in addition to improving Chinese language services, both populations can be satisfied further.

We added a section 7.3 Implications for hotel managers, where we discuss different strategies that managers could follow to improve on customer satisfaction for these populations. We propose two possible strategies, which are appropriate or not depending on a hotel business model. A hotel can aim to increase the numbers of Chinese customers by investing on Chinese language services, which are a larger population but spend less percentage of their budget on hotels. On the other hand, a hotel can aim to increase cleanliness and comfort, which are bigger investments, but would satisfy Western customers, who spend almost half their budget on hotels.

As for the comment "Why do both Westerners and Chinese share dissatisfactory attributes while they show different satisfactory attributes (see Table 6)?" : They are both different. RBO results and our hypothesis testing showed that both satisfactory and dissatisfactory attributes are different for both cultures. Still, I understand that explaining the reasoning behind these results could improve the discussion.

3. Restructuring I found the discussion section (7. Discussion) misrepresents. The section is mainly reporting the findings without in-depth discussion (e.g., discussion on the reason behind such findings). I recommend combining Sections 5 and 6 into one section as an analysis section; and rename 7 Discussion as 6 Results. As suggested in the earlier section, the authors can add a new discussion section with a theoretical discussion. The current 6.1 section can be revised by presenting one-to-one or item-by-item comparison between Chinese and English-speaking customers. The current format is difficult to follow to see which attributes are similar or different between these two groups.

We added most of what was once in the discussion section to the results section, and added more in-depth discussion in section 7 as explained above. The subsections that moved up to the Results section were 6.2 Chinese tourists - A big and clean space, 6.3 Western tourists - A friendly face, and absolutely clean, 6.4 Tobacco, what's that smell?, and 6.5 Location, location, location. The price section was deleted since it was short enough to be discussed elsewhere.

We revised the tables in what previously was 6.1 to present them side to side for better following the differences in keyword ranking. They are now tables 5 and 6.

Reviewer #3 Response

Thank you for your review observations. I will quote each observation and respond to it while indicating my corrections in the text. Aside from these corrections, we also performed a general grammar check and proofread to the new manuscript, including unchanged sections. Because of comments by reviewers, we redid the experiment with more balanced data, so the results shown are different in their quantitative values. However, our conclusions and hypothesis testing remained largely unchanged by this. We also added a new research rationale and focus, which meant new hypotheses were added to the manuscript. The new manuscript focuses on how these populations interact with the high-standard spirit of Japanese hospitality and service called Omotenashi. In this environment, will the cultural differences be strong enough that their satisfaction stems from different factors?, or will both populations be largely satisfied by service and staff related terms?

#1 - Research rationale The authors mentioned "Up until recently, most studies in tourist behavior have been biased for the Western world". Yet, to the best of my knowledge, there are fruitful studies examined Chinese consumers' hotel booking behavior

I searched and found that several similar studies have already been published recently like: * Francesco, G., & Roberta, G. (2019). Cross-country analysis of perception and emphasis of hotel attributes. *Tourism Management*, 74, 24-42. * Huang, J. (2017). The dining experience of Beijing Roast Duck: A comparative study of the Chinese and English online consumer reviews. *International Journal of Hospitality Management*, 66, 117-129. * Jia, S. S. (2020). Motivation and satisfaction of Chinese and US tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management*, 78, 104071.

The authors are encouraged to review and indicate the existence of the above articles. To highlight the uniqueness of this work, the authors can highlight how the current study differs from other published works in terms of "research focus", "methodological approach", "study context" or/and "data size"

I think the misunderstanding comes from the phrasing of our sentence. "Up until recently" does not mean "Until now". We mean to say that in the last recent years the focus has changed. We changed the phrase to reflect this in a more understandable manner. However, we are glad to add these articles to our manuscript literature review and highlight the differences.

In addition, we found a new rationale for our manuscript that makes it stand against the rest: the study context of Japanese hospitality (omotenashi). The different cultures react differently to this high standard hospitality environment, which is interesting in its own regard. We added a new section 3.1 Japanese hospitality: Omotenashi to explain it and developed hypotheses from that con-

text.

We added a paragraph in the beginning of our introduction, modified the abstract, and added the literature review in this context in previously 3.2, now section 3.3 Chinese and Western tourist behavior. We also changed the title to emphasize this context as well.

Because of the change in rationale, new hypotheses were added as well. The main problem in our refocused paper is: In such a high-standard hospitality environment, will the cultural backgrounds react similarly or differently to them? Will they both be satisfied more with hospitality and service related topics than with other topics, or will they ignore this environment?

We found that Chinese tourists don't react as positively to this hospitality as Westerners did. We have two theories which we discussed in section 7. One is that they don't have the same expectations. Getting what they don't expect leads to less satisfaction than if they were expecting it. Another theory is that the language barrier poses an obstacle to the enjoyment of this hospitality.

#2 - Research objectives * Section 2 shows that the objective of this study is to determine the difference in preferences between Chinese- and English-speaking customers of Japanese hotels using text-mining techniques. Yet, according to my review, this study solely examines what English- and Chinese-speaking customers concern most when they evaluate Japanese hotels

* I can accept the assertion that the findings can reflect the differences in factors driving satisfaction and dissatisfaction between Chinese- and English-speaking customers. But the findings cannot reflect their preferences

In a previous version of our manuscript, we had another reviewer point out that "satisfaction and dissatisfaction factors" was a long way to say preferences, which is why we had it corrected this way. Based on this more recent review, however, we happily reverted it to the previous terminology. This is reflected in the manuscript and title. We also modified the research objective to match the new research rationale that was encouraged in a previous point.

#3 - Literature review * Since H1, H2 and H3 has already been extensively tested in prior studies, I do not see the worthiness and necessity to re-examine them.

* For Section 3.2, the authors may consider reviewing more cross-country studies and presenting the findings in a tabular format (with the column showing study topic, industry, samples' nationality ...) This comparison table shall be helpful for the authors to justify if the research gaps truly exist

As to the first point, we believe there is always a necessity in science to re-

examine and avoid making assumptions. In addition to this thought, other reviewers have found the hypotheses helpful in the flow of the manuscript, so we decide to keep them as part of the manuscript. As part of the rationale change, we also added new hypotheses.

For the second point, we agree that adding literature was necessary as per a previous point. We also agree that displaying them such a table would give a helpful overview of the current research in this topic for our paper and have included it in the previously 3.2, now section 3.3.

#4 - Method * To provide a fair comparison, the two sets of texts should review the same hotels within a similar time period. The authors should provide this background information in order to demonstrate the fairness of comparing those two sets of texts

* Please report more information about the review samples, including time period, hotel and hotel attributes involved in this study and etc

In regards to this point, we apologize and would like to explain. In the original manuscript, the experiments were done while the data was still being collected further. After finishing the manuscript, the original data had changed. Not only this, but we realized our mistake from this and another comment from a different reviewer. Indeed the data should reflect the same kinds of hotels, the same kinds of categories, the same time period and attributes. Since we did not data on the attributes of each hotel, we decided to match both English and Chinese comments to come from exactly the same hotels, matching by name. This reduced the number of comments on each language, but it made the data fairer. We don't have data on what type of hotel or star rating they had since this data wasn't available to us, but we can be sure they are the same 580 hotels for both kinds of reviews.

Because of this change in datasets, we also redid the entire experiment, calculating the frequencies and RBO again accordingly.

The time period resulted to be from July 2014 to July 2017. The price range was from 2000 yen per night, to 188,000 yen a night.

We added this paragraph to the Data collection section:

"However, in order to make the data and comparisons we draw from each of these datasets fair, we filtered both databases to only contain reviews from hotels that were in both datasets, using their English names to do a search match. We also filtered them to be in the same date range and cut off reviews outside of each other's date ranges. After filtering, we found that the number of hotels in common in the data collected was 580, and that the common date range for reviews was from July 2014 to July 2017. We also found that the price for a night in these hotels ranges from low priced capsule hotels at 2000 yen per night, to high end hotels 188,000 yen a night as the far ends of the bell curve.

Within this range, from *Ctrip* there was 49,561 reviews comprised of 105,076 sentences, and from *TripAdvisor* there was 41,611 reviews comprised of 352,022 sentences.

We found that after filtering the data, the number of reviews was similar for both English and Chinese reviews, but that English reviews tend to be longer in general.”

#5 - Discussions and Conclusions * Regarding the discussion part, the authors discussed the differences (in sections 7.1 and 7.2) and the similarities (in section 7.3 and 7.4) of Western and Chinese consumers’ satisfaction and dissatisfaction factors. However, the authors discussed the difference in section 7.5 again. Also, the smell of Tobacco was already mentioned in the section 7.2 (“the most common complaint regarding cleanliness was about the carpet, followed by complaints about stains, and cigarette stench in the curtains”). This is a little bit confusing, and I kindly suggest the authors to re-organize the discussion part

* The manuscript does not include much discussions about its theoretical contributions. Again, how what new knowledge was brought by this study? Any contrasting findings were found?

Because of our change in research rationale, we added new content to the introduction, changed the abstract and titles, as well as adding a new section in the discussion with theoretical perspectives on the context of Japanese hospitality. It is right that our study did not measure “culture”, so we can’t be sure of the explanations to certain aspects, although we do explain in the literature review that it has been previously studied how different backgrounds and cultures result in different expectations and as such, different satisfaction factors.

In summary, our expanded discussion is about the differences between Chinese and Western tourists in the context of Japanese hospitality. Western tourists had hospitality and service related satisfaction keywords at the top of their priorities. However, Chinese tourists did not. We theorize this can be from two different causes. One theory is that the Chinese customers have different expectations based on their culture background, and that these expectations lead to different reactions. Specifically, not expecting something and then getting it could be less satisfying than getting an expectation met. Another theory is that despite being in the environment of Omotenashi, their language barriers are enough to disregard hospitality because of their unease or disappointment with the staff and communication availability.

We also expanded on the discussion section by adding more to the previously 7.5, now section 7.2 Managerial vs. Environmental satisfaction, as well as add reasoning to each discussion section as well. In this section we added potential ramifications of the distribution of these factors. While Westerners can be satisfied even more by improving on managerial factors, a big portion of Chinese

satisfaction comes from environmental factors. This means that it could be harder for managers to achieve an improvement in their satisfaction. However, we also claim that improving on cleanliness and making hotels tobacco-free could be a solution that benefits both populations. If this is done in addition to improving Chinese language services, both populations can be satisfied further.

We added a section 7.3 Implications for hotel managers, where we discuss different strategies that managers could follow to improve on customer satisfaction for these populations. We propose two possible strategies, which are appropriate or not depending on a hotel business model. A hotel can aim to increase the numbers of Chinese customers by investing on Chinese language services, which are a larger population but spend less percentage of their budget on hotels. On the other hand, a hotel can aim to increase cleanliness and comfort, which are bigger investments, but would satisfy Western customers, who spend almost half their budget on hotels.

We added most of what was once in the discussion section to the results section, and added more in-depth discussion in section 7 as explained above. The subsections that moved up to the Results section were 6.2 Chinese tourists - A big and clean space, 6.3 Western tourists - A friendly face, and absolutely clean, 6.4 Tobacco, what's that smell?, and 6.5 Location, location, location. The price section was deleted since it was short enough to be discussed elsewhere.

We also revised the tables in what previously was 6.1 to present them side to side for better following the differences in keyword ranking. They are now tables 5 and 6.

#6 - Minor * The manuscript title should be revised. The phrase "drivers of satisfaction and dissatisfaction" should be added. Also, instead of mentioning "with machine learning" in the manuscript title, the authors may consider highlighting "using Rank-biased Overlap measure"

We reconsidered the title from this and another point in this review as follows:

"Cross-culture differences in tourists faced with Japanese hospitality: A text mining and rank-biased overlap measure study of satisfaction and dissatisfaction factors in Chinese and Western cultures"

This reflects both points in this comment and focuses on the change in research rationale, while also highlighting the method.

Reviewer #5 Response

Thank you for your review observations. I will quote each observation and respond to it while indicating my corrections in the text. Aside from these corrections, we also performed a general grammar check and proofread to the new manuscript, including unchanged sections. Because of comments by reviewers, we redid the experiment with more balanced data, so the results shown are different in their quantitative values. However, our conclusions and hypothesis testing remained largely unchanged by this. We also added a new research rationale and focus, which meant new hypotheses were added to the manuscript. The new manuscript focuses on how these populations interact with the high-standard spirit of Japanese hospitality and service called Omotenashi. In this environment, will the cultural differences be strong enough that their satisfaction stems from different factors?, or will both populations be largely satisfied by service and staff related terms?

This study used Ctrip as the data base. However, majority users of Ctrip is Mainland Chinese, but Chinese-speaking tourists can also be Taiwanese, Malaysian Chinese, Singaporean Chinese and Hong Konger too. Therefore I suggest the authors change the term Chinese-speaking tourists to Mainland Chinese tourists.

Thanks for pointing this information out. We pointed this out in the text.

In literature review, the authors listed three hypothesis relating to Chinese and English speaking tourists in 3.1 and then a separate section 3.2 to explain the difference between Chinese and Western tourists. I would suggest the authors combine these two sections so as to give a more comprehensive view on the setting of the three hypothesis.

We respectfully disagree with this suggestion since the content of section 3.1 and 3.2 are very different. While 3.1 sets the premise of differences in customer satisfaction and the reasoning behind them existing, 3.2 discusses previous literature on specific behaviors and cultures of each kind of tourist. Since we added a new section 3.1, these two sections are now 3.2 and 3.3 respectively.

In formular (1), what are alpha, x and y?

We added a description of the formula to the manuscript, as well as a short description of the SVM algorithm to make it more understandable.

In the Data Collection, can the authors explain what kinds of hotels (category, Hotel types, or star rating) were extracted? Are the same hotels extracted from both Ctrip and Tripadvisor? How many hotels are there? I believe different types of hotels would have generated different satisfaction level and perspectives. Please

provide more details about these hotels.

In regards to this point, we apologize and would like to explain. In the original manuscript, the experiments were done while the data was still being collected further. After finishing the manuscript, the original data had changed. Not only this, but we realized our mistake from this and another comment from a different reviewer. Indeed the data should reflect the same kinds of hotels, the same kinds of categories. Since we did not have these data, we decided to match both English and Chinese comments to come from exactly the same hotels, matching by name. This reduced the number of comments on each language, but it made the data fairer. We don't have data on what type of hotel or star rating they had since this data wasn't available to us, but we can be sure they are the same 580 hotels for both kinds of reviews.

Because of this change in datasets, we also redid the entire experiment, calculating the frequencies and RBO again accordingly.

The time period resulted to be from July 2014 to July 2017. The price range was from 2000 yen per night, to 188,000 yen a night.

We added this paragraph to the Data collection section:

"However, in order to make the data and comparisons we draw from each of these datasets fair, we filtered both databases to only contain reviews from hotels that were in both datasets, using their English names to do a search match. We also filtered them to be in the same date range and cut off reviews outside of each other's date ranges. After filtering, we found that the number of hotels in common in the data collected was 580, and that the common date range for reviews was from July 2014 to July 2017. We also found that the price for a night in these hotels ranges from low priced capsule hotels at 2000 yen per night, to high end hotels 188,000 yen a night as the far ends of the bell curve. Within this range, from *Ctrip* there was 49,561 reviews comprised of 105,076 sentences, and from *TripAdvisor* there was 41,611 reviews comprised of 352,022 sentences.

We found that after filtering the data, the number of reviews was similar for both English and Chinese reviews, but that English reviews tend to be longer in general."

In Table 2, I am quite surprise to see that the last item "Chinese person" can be the top keywords with only 16 occurrences. Can the authors explain why there are so few negative Chinese keywords?

Since we repeated the experiment, the frequencies have changed, however this keyword still only has 16 occurrences. The lack of negative keywords comes from a peculiarity of the Chinese reviews from Ctrip. Most of the reviews point out the positive extensively, while only a few sentences are negative. It's a limitation of our paper that the negative keywords depend on the sample taken

at the beginning, since we could not label more manually at the time.

Regarding the keyword "Big", this is a very general word and can be describe as "Big" in size, "Big in feelings, or could be negative too "big disappointment". Can the authors explain further how this can be avoided in your data analysis?

One thing to note is that using SVM, we labeled sentences as positive or negative, so negative usage of the word when counting the frequencies is negligible. In addition to this, we also had the same question when we saw our results the first time, and took a small sample to observe the usage of the word. However we didn't deem it appropriate in the text at first, but with this comment and thinking it over once again, we have added the next paragraph to the manuscript:

“From these two first ranking keywords (“big” and “clean”), we can assert that room quality is the most critical satisfaction factor for Chinese customers. However, the word “big” is widespread in the language and could be used in different contexts to point out positive and negative aspects. Since we are interested in how the word is a keyword of positive sentences, we took a small sample of the positively classified sentences that contained the word “big” in Chinese. A team of Chinese speaking lab members then pointed out the usage of the word. Generally speaking, it was used for the spaciousness of rooms, bathtubs, and communal bath facilities.”

For the managerial and environmental keywords, what criteria were used for classifying these keywords?

We explain the difference between managerial and environmental factors in different sections of the manuscript. However, since it might still be unclear, we specified the method for classifying the keywords in the Results section, adding the following text:

“We manually labeled the top 625 keywords of each language into either managerial or environmental by considering how the word would be used when writing a review. If the word implies an issue that could be solved or managed by the hotel staff or management, we consider it managerial. If the word is describing factors that are unchangeable by the staff or management, we consider them environmental. The interpretation of these keywords is shown in the Tables 8 and 9.”

We also added the tables that were previously in the appendix clearing this issue into the main text for readability.

Editorial issues The Appendix is actually part of the manuscript. This manuscript contains too many Tables. First some was mentioned in the manuscript but appeared in Appendix. For example line 306, Table 7 and Section B was mentioned. What is Section

B? Please move all relevant tables to the manuscript but not in Appendix. I would suggest the authors to consider to either combine some of them or not to mention them in the manuscript. Besides, the numbering of the tables was not in sequence in the manuscript. Line 360, after Table 3 and 4, it suddenly jumps to Table 10.

We moved some tables that we deemed important out of the appendix, but other tables got removed as per this request. We also joined the tables with the top keyword results as per this and another reviewer's request.