Research on Consumer Choice Behavior by Reviews on Expedia

Xingchen Zhou

Woodsworth College University of Toronto (U of T) Toronto, Canada, M5S 1A9
xchen.zhou@mail.utoronto.ca

Abstract

Expedia is a travel searching and booking platform. The dataset specifically showcases property searches on the platform. To help Expedia better know how to attract more consumers, by studying the association between reviews and consumers' choices by using hypothesis test and multiple linear regressions. The results of the study found that reviews can significantly affect consumers' choice behavior. Secondly, consumers tend to view properties with Review counts less than 5000 and with higher Average Guest Rating. Thirdly, consumers prefer to give higher individual ratings to properties with more Review Count. Lastly, consumers tend to choose properties with Star Rating at 4, and properties with higher Star Ratings tend to have obviously more Review Counts. The research results of this paper can make some policy suggestions for managers engaged in this field and have important practical significance in order to regulate the healthy development of the platform.

Keywords-*Expedia, Property, Online Platform, Review*

1. INTRODUCTION

1.1. Research Background and Motivation

Recently, it is a trend that consumers use online resources more for tourism and hospitality options [1]. Consumers normally will make decisions based on after learning others' experience, and pervious consumers' experience can be greatly recorded in the form of reviews nowadays on the e-commercial platforms. These reviews can be regarded as useful references for potential consumers [2]. Nowadays, there exists many kinds of reviews, such as guest rating scores, agency rating scores and textual reviews. Across diverse social media, the role of online review platforms in tourist decision making is particularly noteworthy [3]. Consumer reviews for online platforms can influence consumers' purchase decisions [4]. Properties on the platform of Expedia can be regarded as experiential products, since the quality of properties can be determined after consumption. During the process of purchasing, the special characteristic makes consumers tend to depend more on others' recommendations [5-7]. Therefore, it is important to manage the reviews for every property online. Firstly, consumers' decisions will rely on seeking some information from reviews, because the properties are experiential products. Secondly, properties as experiential products, which means consumers take

many involvements which are incentives for consumers to generate reviews [8]. Thirdly, properties might be influenced by reviews more, because the special features for example, lifeline of hotels is much longer.

Since consumers will get more information before making final purchasing decisions, to reduce risks. Consumers may need many intrinsic and extrinsic cues for references. Therefore, this is a good point for managers to focus on improving, after knowing the influential relations between the reviews and consumers' preferences and choices. This research focuses on examining how reviews will influence consumers' decisions.

1.2. Literature Review

The research result related to the relation between the reviews and consumers' choices, from the aspects of Average Guest Rating, Star Rating and Review Count. Reviews can be categorized into two groups, one is reviews from consumers, normally, this is kind of reviews includes the data such as individual guests' reviews, average guests' reviews and review contents, from online platforms. Another one is reviews from relevant agencies, such as tourism, and this is normally displayed as the star rating.

1.2.1. Online Review

Reviews can play similar roles as word-of-mouth in some extend, which is an important source of influencing consumer behaviours. As a consumer, it seems that other consumers' reviews are sometimes more believable. Due to the uncertainties, consumers tend to refer others' opinions or experiences before consumption. Zhu and Zhang (2010) also point out that for less popular products, online reviews seems to be more influential, and online reviews are comparatively influential for products after experiential people used [4]. Moreover, apart from statistics, customers read review text to catch information, consumers incline to give more positive reviews if they focus on specific experiential features [9]. Meanwhile, positive reviews can increase consumers' trust of online products. In some extend, online reviews can be regarded as consumers' behavior from the aspects of information searching and sharing, which can bring benefits if managers in tourism industry effectively use this [2]. Besides, a larger number of reviews may increase the objectivity and be trusted more by potential consumers [10]. The review ratings can reflect the overall quality of properties based on many different consumers' preferences [11]. Since consumers take reviews as a relatively credible source of information, the rating of reviews has a strong effect on consumer judgement [12]. Building on the findings of Balakrishnan and Worndl (2021), this paper finds that end users of travel recommender systems largely rely on the help of recommender systems when dealing with external non-functional influences [13]. This means a consumer will find other information from product-related resources to decrease the uncertainty. Before consumers make their final purchase decision, consumers will get information cues from other resources first, then they will adjust the perceived quality of the product based on this [10]. Multiple dimensions of products will influence the consumers' expectations on the products [14].

1.2.2. Star Rating

Properties' star rating can be regarded as a general classification standard of properties' quality. In different countries, the star rating might be controlled by different agencies. For example, in Europe and some parts of Asia, star rating is set by popular and influential tourism institutions [10]. However, if some properties do not have star ratings, some popular travel websites also give a reliable rating to determine the properties' quality. The star rating is a stable signal for a period comparing to other factors, and it can reflect the quality [15]. The star rating is another way to increase consumers' cognitions on properties' quality, and it can reflect the reputations of products in the tourism industry. In some extend, star rating might change the influence of reviews. Facing properties with lower star rating, consumers tend to depend more on reviews' information [10]. By contrast,

reviews will be less influential if the property has higher star rating.

1.3. Research Contents and Framework

The objective of this research is to investigate the moderating effect of reviews including Review Count, Review Rating and Star Rating, on consumers' choices. The structure of this paper as follows: the first part is about the background and literature reviews. The second part focuses on data and analyzing methods. In part 3, we post this research's results and managerial implications, and list some problems valuable to discuss. Lastly, we give the conclusion for the whole research.

2. METHODOLOGY

2.1. Data Description

The data for this study is collected from the RecTour community at the 15th ACM Conference on Recommender Systems released by the Expedia Group [16]. Expedia is a travel searching and booking platform, which is a leading in the field of online travel service and operates among many countries [10]. Therefore, using data collected from Expedia can provide a better understanding of the overall online booking platforms of properties, since properties and data from Expedia include a wide range over the world. The review system of Expedia is special that only consumers who finished their trip can be allowed to give reviews [10].

The dataset collected data from 2021-06-01 to 2021-07-31 and contain searches for a random sample of consumers who made at least one click during the above time frame. This data consists of global lodging shopping and purchase data from consumers in multiple countries across tens of thousands of destinations [16]. The data are organized around a set of "search result impressions", i.e., the ordered list of properties that a consumer sees after a lodging search on the Expedia sites [16]. The user response is provided as a click on a property or/and a purchase of a property room, with only clicks and purchases that occurred after a search and before the next search within a 180-minute time limit are attributed to a search [16]. This data specifically showcases property searches on the platform and some property amenities of properties. "Property" refers to hotels, vacation rentals, apartments, B&Bs, hostels or other rental-housing options appearing on the brand Expedia's websites [16]. The room types are not distinguished, and the data can be assumed to apply to the least expensive room type [16]. If a consumer books 4 distinct properties, the data is excluded. The data span more than 800k unique users and approx. 25M searches and include desktop and mobile device traffic [16]. The data include traveler inputs such as adding filters and selecting specific sort types, such as price ascending [16]. Data in this dataset is anonymous,

by taking several steps to anonymize and obfuscate the true data distribution to protect users and commercial sensitivities [16]. But the data provider changed the proportions of the number of clicks and the rate of transaction. Many previous research study this field by using consumers' reactions as the real response, such as giving reviews or not, because there exist time differences between finishing transactions and data collections. In order to solve the uncertainty of this dataset, though it is provided based on real traveler behaviors, we choose to use the number of clicks as the representatives of the transactions [16].

To be noted, based on previous research, recent reviews can be more influential than reviews before, besides, before making decisions, most people only focus on the first page [17].
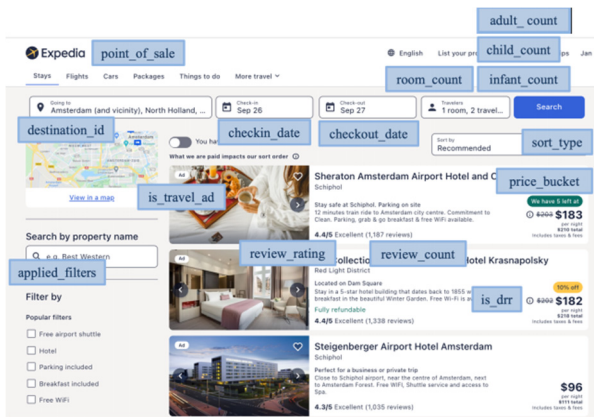

Figure 1. Data labels as seen on Brand Expedia website

Figure 1 outlines the relationship between the search and property data in the dataset with the values impressed on the Brand Expedia website. Combine click and purchase paths on the Brand Expedia website. From Figure 1, we can find that there are four kinds of review elements, including the individual guest rating, the average guest rating, the star rating and the total number of reviews. The guest rating can demonstrate the satisfactory of consumers, because it is gave after consumers' stay. The star rating is a longitude data instead of ratings from consumers in time [10]. Besides, the star rating can reflect the quality of properties. The review counts displayed for all properties establish the overall popularity of this property.

2.2. Data Analysis

2.2.1. Basic Description

Based on prior studies, we assume that when a customer visits a property, he/she will read other customers' reviews, and apply the reviews' ratings for booking reference. Using summary statistics including mean values, standard deviations, minimums and maximums to know detailed information for the average Expedia guests ratings (Average Guest Rating), the total

number of reviews (Review Count), the star rating (Star Rating), and the number of clicks of one property (Click Count). The descriptive statistics of the properties' information are shown in Table 1.

TABLE 1.   KEY DESCRIPTIVE STATISTICS

Variable	Mean	Standard Variance	Minimum	Maximum
Average Guest Rating	4.11	0.80	0	5
Review Count	1143.99	2152.44	0	25225
Star Rating	3.72	0.87	0	5
Click Count	0.06	0.26	0	4

2.2.2. Basic Analysis

To know the relation between reviews and the reactions of consumers, from the perspectives of click counts and average guest rating.
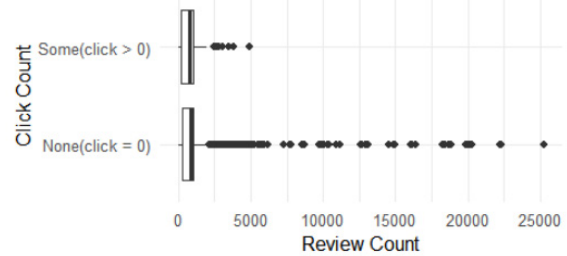

Figure 2. The distribution of Click Count with Review Count

From Figure 2 show that the distribution of review count and click count. This article find that there are exists many outliers for reviews counts with clicks or without clicks, which are not so important in our research. But it seems that there only exist small differences for the medians between review counts with clicks or without clicks.
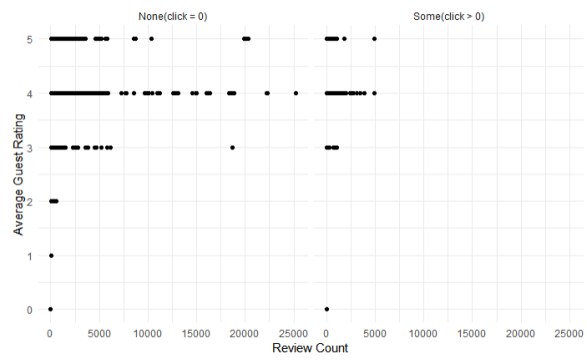

Figure 3. The distribution of Average Guest Rating with Review Count under Click Count

Figure 3 show that an association that more review counts have higher average guest rating mostly. It is obvious that properties with less review counts have spread average guest rating, with some properties in higher average guest rating. Besides, consumers tend to view properties with review counts less than 5000. Based on the review system of Expedia that only guests who consume can give the individual ratings and reviews, and we regard every review as a consumer. In some extend, we can say that consumers tend to choose properties with higher Average Guest Rating.
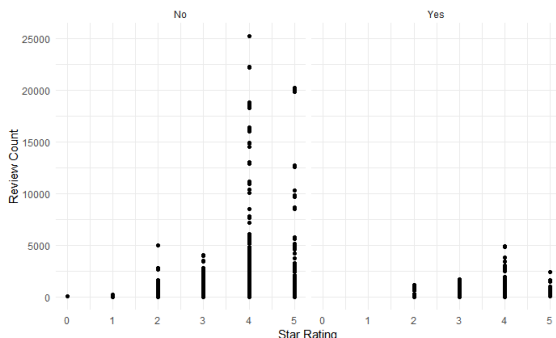

Figure 4. The distribution of Star Rating with Review Count under Click Count

As the description of the dataset, the Star Ratings are ratings coming from agencies, which are stable in a short period. From the Figure 4, this paper can find that normally, properties with higher Star Ratings will have obviously more Review Counts. Meanwhile, consumers prefer to choose properties with Star Rating at 4 instead of the highest rating at 5 or lower rating like 3.

2.3. Hypothesis Test

From the Figure 2 and Figure 3, this article find Review Count will influence the Average Guest Rating, but it seems that the Review Count has no influence on the Click Count. In order to check this, we conduct the Hypothesis Test. Because there are exists many outliers which are influential but not the research's focus, so we choose to use the median values to decrease the influence of outliers. Setting the Null Hypothesis as the median value of review counts with no click should be almost equal to the median value of review counts with clicks.

H_0: Median$_{some}$ − Median$_{none}$ = 0

Setting the Alternative Hypothesis as the median value of review counts with no click should not equal to the median value of review counts with clicks.

H_1: Median$_{some}$ − Median$_{none}$ ≠ 0

After calculating a test statistic, setting this as $\hat{p}$ = −125. Simulate samples under the null hypothesis H0 for 1000 times and calculate the statistic for each one. Then, evaluate the evidence against H0 by calculating the p-value. The result of P-value is 0.395, which is bigger than

0.1. Thus, we have no evidence to reject the H0. Therefore, from this Hypothesis test we can conclude that the review counts have almost no influence on whether there exist clicks or not.

However, generally, information cues about products influence consumer purchase decisions, and prior studies show that when purchasing experience goods, consumers incline to depend more on the recommendations of others, versus search goods [18].

2.4. Multiple Linear Regression

From the basic analysis above, we can find Star Rating has some associations with the Review Counts and Click Count. However, the result about no association between the Click Count and Review Count is out of the expectations. To find whether there will be a relation between Click Counts and Review Count, under some other factors, such as Average Guest Rating and Star Rating, we choose to use multiple linear regression to test.

TABLE 2. RESULTS OF USING THE MULTIPLE LINEAR REGRESSION

Variable	Signif.
(Intercept)	0.345
Review Count	0.799
Review Rating	0.967
Star Rating	0.572
Review Rating * Review Count	0.917
Review Count * Star Rating	0.706
Review Rating * Star Rating	0.819
Click Count * Review Rating* Star Rating	0.802

From Table2, it can find that there exist an association among Click Counts, Review Counts, Average Guest Rating and Star Rating.

3. RESULTS AND DISCUSSION

3.1. Results

From the analysis above, we can find several results and based on these results, we can list corresponding suggestions to Expedia platform, which might also be suitable. Firstly, reviews will influence the choices of consumers. Although Review Count has no direct contribution on influencing consumers' choices to click, the total reviews including Review Count, Average Guest Rating and Star Rating, have moderating effect on consumers' clicks. This demonstrates that consumers will make their final choices or determinations after knowing more details about properties from several directions. Secondly, consumers tend to view properties with Review counts less than 5000 and with higher Average Guest Rating. Thirdly, consumers prefer to give higher individual ratings to properties with more Review Count. Individual ratings are the basic combinations of the Average Guest Rating, and it can reflect consumers'

satisfactions to the properties in time. Based on the review system of Expedia that only guests who consume can give the individual ratings and reviews, we can find a phenomenon that previous good properties will be better or remain good continually, while properties with lower ratings might be hard to change its popularity. Fourthly, consumers tend to choose properties with Star Rating at 4, and properties with higher Star Ratings tend to have obviously more Review Counts. Star rating is a stable statistic in a short period, which is set by agencies. From the reactions of consumers, we can regard Star Rating as a represent of the quality for a property in a large extend.

Based on the results above, managers should focus on the importance of customers' reviews, reflecting on the reviews, and encourage them to post reviews, which can help new consumer better get to know the properties. Meanwhile, more reviews can create a reliable feeling to consumers that it is reliable and valuable to choose such properties, as many people have already tried before with good experience. With more reviews means the property is more popular, so this property might have higher Star Rating from agencies, but it is better to keep the Star Rating around 4, which is the most popular stage that consumers will choose.

3.2. Discussion

From the results, we find that consumers tend to view properties with Review Count less than 5000 but prefer to choose properties with have higher Review Rating or higher Star Rating which can be improved by having higher Review Count. This is a conflict situation. In this research, we use the total number of reviews as the Review Count to analyze, but we do not pay attention to the textual part of reviews. The reasons to cause this conflict situation might be the textual influences from reviews. There is a standpoint that if hotels receive positive reviews, room price and star rating can cause less influences on their online sales performance, but for receiving negative reviews, the online sales performance seems to be influenced by room price and star rating more [10]. Therefore, there still need more research on how textual reviews influence the property bookings and viewings.

Meanwhile, the data we used to determine consumers' final transactions is based on the distributions and number of Review Count. Even though based on the review system of Expedia that only consumers transacted on the platform can give reviews, which means Review Count can represent as the booking data, there still exits some differences between the real data and this estimated one. Since Review Count is a long-term accumulation of data, it can lay the foundation for this article to understand more consumers' preferences in a timely manner.

4. CONCLUSION

The objective of this research is to investigate the relation between reviews and consumers' choices. We use the data from the platform of Expedia, mainly focusing on the Review Count, Review Rating, Star Rating and Click Count. Methods used in this research contain Hypothesis Test and Multiple Linear Regression. Review Count and Review Rating can reflect the recent situations in some extend, while Star Rating is a stable statistic during a short term, which can used to reflect the change and influence during a period. Based on previous research, we find there are exists relation between reviews and consumers' choices, but in our research, we use latest data to generate a more detailed analysis. The results reveal that firstly, reviews have moderating effect on the choices of consumers, but Review Count has no direct contribution on influencing consumers' choices to click. This demonstrates that consumers will make their final choices or determinations after knowing more details about properties from several directions. Secondly, consumers tend to view properties with Review counts less than 5000 and with higher Average Guest Rating. Thirdly, consumers prefer to give higher individual ratings to properties with more Review Count, which are the basic combinations of the Average Guest Rating, and it can reflect consumers' satisfactions to the properties in time. Fourthly, consumers tend to choose properties with Star Rating at 4, and properties with higher Star Ratings tend to have obviously more Review Counts.

The findings from this study can bring many benefits for both researchers and practitioners. Based on the results above, we list some suggestions to managers in this field. Firstly, managers should emphasize the importance of customers' reviews, from the aspects of the number of reviews and the quality of reviews. Secondly, raise the Star Rating to around 4, which is the most popular stage that consumers will choose. The paper concludes by addressing the limitations of our study and discussing future research directions. Firstly, we shall pay attention to the textual part of reviews. Whether the contents of the reviews are positive or negative, might also influence consumers' viewing and booking. Secondly, we shall collect more actual and immediate dataset to study, since the data we used to determine consumers' final transactions is based on the distributions and number of Review Count. However, the Review Count is an accumulated data for a long time, but we want to know more about consumers' preference in time.

REFERENCES

[1] A. R. Alaei, S. Becken, & B. Stantic. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research, 58*(2), 175–191.

[2] J. A. Chevalier, & D. Mayzlin. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research, 43*(3), 345–354.

[3] T. W. Lui, M. Bartosiak, G. Piccoli, & V. Sadhya. (2018). Online review response strategy and its effects on competitive performance. *Tourism Management (1982), 67*, 180–190.

[4] F. Zhu, & X. Zhang. (2010). Impact of online consumer reviews on Sales: The moderating role of product and consumer characteristics. *Journal of Marketing, 74*(2), 133–148.

[5] C. Park, & T. M. Lee. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research, 62*(1), 61–67.

[6] S. Senecal, & J. Nantel. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing, 80*(2), 159–169.

[7] D. Weathers, S. Sharma, & S. L. Wood. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. *Journal of Retailing, 83*(4), 393–401.

[8] D. Stokes, & W. Lomax. (2002). Taking control of word of mouth marketing: The case of an entrepreneurial hotelier. *Journal of Small Business and Enterprise Development, 9*(4), 349–357.

[9] J. Joško Brakus, B. H. Schmitt, & S. Zhang. (2014). Experiential product attributes and preferences for new products: The role of processing fluency. *Journal of Business Research, 67*(11), 2291–2298.

[10] M. Wang, Q. Lu, R. T. Chi, & W. Shi. (2015). How word-of-mouth moderates room price and hotel stars for online hotel booking: An empirical investigation with Expedia data. *Journal of Electronic Commerce Reseasrch, 16*(1), 72-80.

[11] W. Duan, B. Gu, & A. B. Whinston. (2008). Do online reviews matter? — An empirical investigation of panel data. *DECISION SUPPORT SYSTEMS, 45*(4), 1007–1016.

[12] R. Grewal, T. W. Cline, & A. Davies. (2003). Early-Entrant Advantage, Word-of-Mouth Communication, Brand Similarity, and the Consumer Decision-Making Process. *Journal of Consumer Psychology, 13*(3), 187–197.

[13] G. Balakrishnan, & W. Worndl. (2021). Multistakeholder Recommender Systems in Tourism.

[14] D. Grewal. (1995). Product Quality Expectations: Towards an Understanding of Their Antecedents and Consequences. *Journal of Business and Psychology, 9*(3), 225–240.

[15] A. A. Israeli. (2002). Star rating and corporate affiliation: their influence on room price and performance of hotels in Israel. *International Journal of Hospitality Management, 21*(4), 405–424.

[16] A. Woznica, & J. Krasnodebski. (2021). Expedia Group RecTour Research Dataset.

[17] P. A. Pavlou, & A. Dimoka. (2006). The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Buildings, Price Premiums, and Seller Differentiation. *Information Systems Research, 17*(4), 392-414.

[18] C. Park, & T. M. Lee. (2009). Information Direction, Website Reputation and Ewom Effect: A Moderating Role of Product Type. *Journal of Business Research, 62*(1), 61-67.