Chapter 9

UNDERSTANDING GUESTS' DISSATISFACTION Application of Data Visualization Tools

Cindy Yoonjoung Heo

University of Applied Sciences and Arts Western Switzerland, Switzerland

Bona Kim

Singapore Institute of Technology, Singapore

Laetitia Drapé-Frisch

University of Applied Sciences and Arts Western Switzerland, Switzerland

Abstract: Organizations often view current market situations from the customers' viewpoints. To do so, they face an incredibly vast volume and variety of data. Data visualization makes huge amounts more accessible and understandable and helps to communicate complex information more accurately and effectively. This chapter discusses how to utilize user-generated data in the tourism industry to enhance the customer experience through data visualization tools. Five analysis tools are used to identify the factors that contribute to hotel guests' dissatisfaction in five top city destinations. Identifying and eliminating dissatisfaction is the first step in enhancing the quality of the experience and in fostering loyalty in the long run.

Keywords: Data visualization; big data analysis; tourist experience; customer dissatisfaction; hotel industry; five top city destinations

INTRODUCTION

Today's service providers try to make customer-centric decisions to provide a positive experience and to generate repeat business, loyalty, and profits. They view current market situations from the customers' viewpoints rather than from their own business perspectives. Tourists leave a data trail from the moment, even before, they decide to travel until after they finish their trip. In particular, as social media on the Internet has grown dramatically (Browning, So, & Sparks, 2013; Xiang, Schwartz, Gerdes Jr, & Uysal, 2015), tourism agencies such as destination marketing/management organizations, hotels, and distribution channels are trying to utilize user-generated data to expand their knowledge of target markets. They do so in order to maintain loyal customers and to identify opportunities to attract new ones. Online user-generated reviews are not only an important source of information for all (Pan, MacLaurin, & Crofts, 2007), they are also recognized as a resource enabling service providers to improve the quality of the experience. Studies have found a significant relationship between online consumer reviews and the business performance of hotels (Ye, Law, & Gu, 2009).

However, as the amount of data available is increasing exponentially, organizations easily get confused by their complexity. Big data is generally defined as:

enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios: from scientific computing applications to social networks, from e-government applications to medical information systems, and so forth (Cuzzocrea, Song, & Davis, 2011, p. 101).

These huge volumes of data include various types such as text, picture, video, audio, and weblog. Tourism is a data-rich domain and the industry is one of the core areas where effective use of big data can change the way business is done. Tourism businesses have recognized the importance of data analytics, but they face an incredible volume and variety with which to work.

Data visualization makes huge amounts of data more accessible and helps to communicate complex information more effectively. Visualization assists individuals to highlight the core meanings of big data by displaying and depicting patterns, trends, and correlations (Dimara, Bezerianos, & Dragicevic, 2017). Although various visualization tools and software are

available to users, their application is limited. This chapter presents ways in which different visualization tools can be effectively used to analyze user-generated hotel reviews from Expedia. Specifically, it focuses on negative review comments, because understanding what makes customers unhappy is the first step for hotels to generate more demand and improve their financial performance (Sparks & Browning, 2011). The features and results of five visualization tools were compared to identify customer dissatisfaction factors and to discuss how to utilize user-generated data in the tourism industry.

DATA VISUALIZATION TOOLS

Customers' online reviews play an important role in generating electronic word of mouth, which has a wider reach and creates faster interactions compared with traditional word of mouth (Cantalops & Salvi, 2014). Therefore, scholars and practitioners recognize the importance of analyzing user-generated content on social media to improve the quality of their customer experience. Recognizing significant factors determining satisfaction and dissatisfaction provides crucial opportunities for service recovery, which is ultimately related to the service firms' profitability (Kim, Kim, & Heo, 2016). However, research efforts remain insufficient with regard to fully leveraging customers' online reviews in an effort to understand their experiences (Kim et al., 2016; Levy, Duan, & Boo, 2013; Li, Ye, & Law, 2013).

In general, researchers use text analysis approaches to understand consumer behavior from user-generated content, because statistics-based data analysis methods can be ineffective in extracting information from text-based reviews and comments (Li, Law, Vu, Rong, & Zhao, 2015). Common approaches for the text analysis are to manually analyze the text and make references to relevant insights from the content (Cheng & Edwards, 2019). However, manual methods are time consuming and have a few limitations, not to mention the ever-increasing volume of data (Agrawal, Kadadi, Dai, & Andres, 2015). Massive quantities go unused or underused, because people are not able to find the quantities and relationships within the pool. Finding the valuable information hidden in big data is not an easy task.

Compared to textual formats, data visualizations can highlight relations in the data, facilitate the recognition of patterns, and reduce cognitive load (Dimara et al., 2017). When properties of symbolic data are mapped to visualize a few properties, humans can browse through large amounts efficiently, because they have very well-developed visual skills and capacities.

According to Vogel, Dickson, and Lehman (1986), the presentations using visual support are 43% more persuasive. Presenters who use visual support are reported to be clearer, more concise, professional, persuasive, interesting, and make better use of supporting data.

Many researchers in tourism and hospitality businesses have adopted various visualization tools to explore a range of issues. One of the most commonly used tools is Leximancer adopted in analyzing literature and customer reviews in several studies (Li, Pearce, Low, 2018; Pearce & Wu, 2016; Wu, Wall, & Pearce, 2014). By adopting the unsupervised analysis tool, Li et al. (2018) analyzed 450 media texts to understand the current issue and lead a debate about digital-free tourism and how media representation has been changed in the social context. Wu et al. (2014) examined international tourists' experiences in the most well-known tourist market in Beijing by analyzing reviews on Trip Advisor using the Leximancer tool. Depending on satisfaction levels and gender, different tourists' experiences are visually presented. In another study by Pearce and Wu (2016), tourists' experiences at a small, themed attraction are presented in viewing interacting elements among 167 tourists' reviews. The analysis tool has a unique feature in identifying dominant concepts and phrases in linkages. The adequate reviews and detailed presentation of Leximancer in these papers means that it is not included in this study. Instead other tools which are less well-known are explored.

Another tool is OpinionSeer. Wu et al. (2010) used it in order to visually analyze a large collection of online hotel customer reviews. Bjørkelund, Burnett, and Nørvåg (2012) focused on opinion mining and described how the results of sentiment analysis of textual reviews can be visualized using Google Maps. Li et al. (2015) utilized Emerging Pattern Mining to identify emergent hotel features of interest to international tourists. Several automated approaches, such as corpus-based semantic analysis (Rayson & Garside, 2000) and stance-shift analysis (Davidson & Skinner, 2010), are also employed; but such approaches require researchers to have knowledge of linguistics and access to expensive software (Capriello, Mason, Davis, & Crotts, 2013). Further, research comparing the different visualization tools is rare.

Study Methods

Five data visualization analysis tools—namely, Voyant Cirrus, Lexos—BubbleViz, Voyant Links, KH coder-Co-occurrence network, and Wordwanderer—are applied in this study. Online hotel reviews were

retrieved on Expedia.com, which is one of the most dominant online travel agency websites for both hotel's booking and travel reviews. Five top city destinations based on the number of international tourists in 2014 (Euromonitor International, 2016) were selected for data analysis (Hong Kong, London, Singapore, Bangkok, and Paris). All 5-star hotels in the five destinations (23 in Hong Kong, 97 in London, 41 in Singapore, 47 in Bangkok, and 74 in Paris) were selected to investigate the experiences of luxury hotel customers. All negative reviews (the total of 3,827 reviews which are rated as 1 and 2 out of 5) were collected for analysis using data visualization tools (Table 1).

Results by Voyant Cirrus

This tool provides a word cloud to exhibit a visualized corpus in accordance with frequency of words (Voyant Cirrus, 2012). The more they are mentioned, the more they are placed at the center, and the bigger they are displayed in terms of size. The cloud depicts words in the corpus and the importance of them. Researchers can configure the level of inclusive words ranged from 25–500. However, every time they reload the corpus, words and colors in a word cloud may appear in different ways; in turn, the colors of words and their absolute positions are not important. The frequency of each word also appears by clicking on it.

Table 1. Data Profile

City Destination	Hong Kong (n = 388)	London (n = 1,107)	Singapore (n = 798)	Bangkok (n = 739)	Paris (n = 795)
Total No. of words	25,031	100,895	71,223	64,530	65,780
No. of unique word forms	3,254	6,793	5,597	5,478	5,539
Average Words Per Sentence	17.7	16.9	17.5	17.9	17.0
Top five most frequent words (No. of frequency)	1st room (407)	room (1,579)	room (1,107)	hotel (1,086)	hotel (1,160)
	2nd hotel (402)	hotel (1,525)	hotel (979)	room (794)	room (1,043)
	3rd service (120)	staff (446)	check (304)	staff (312)	star (334)
	4th stay (116)	star (417)	staff (296)	service (257)	staff (326)
	5th staff (110)	service (365)	time (256)	good (236)	service (230)

This tool explores digitized texts in a corpus by creating a large number of bubbles in different colors in a map. The map highlights the importance of each word in an overall picture (Lexos Installer, 2016). In order to create a bubble map, graph size, maximum number of words and minimum word

Note: London dataset; each of the words is highlighted in a variety of colors

length: 3. The bubble map generated from the Singapore dataset presents the words “room” and “hotel” in two large bubbles of dark gray.

Results by Voyant Links

This tool shows the collocation of words in a corpus by representing them in a linkage using a directed graph (Voyant Links, 2012). The main functions for data visualization are the size of the node and the thickness of the edge that depicts how frequently they are mentioned in the corpus and linked in the resulting graph. The outcome table consists of further information in the form of a summary table including total number of corpus, number of unique word forms, average number of words per sentence, and most frequent words in a corpus, as well as the frequency at which words appear. For example, the data in the corpus of Paris 5-star hotels includes 65,780 total words and 5,539 unique word forms. The most frequent words in the corpus are hotel (1,160), room (1,043), star (334), staff (326), and service (230).

This tool provides a highlighting function that depicts links with a word (highlighted in dark gray) when the particular one is identified. For example, when the most frequently mentioned word “hotel” is selected, all the linkages connected with “room” are highlighted in dark gray. The main strength is that researchers can experiment with diverse visualized outcomes by adjusting the level of context ranging from 3 to 30. Figure 3 shows visualized linkages between keywords at the 5 and 15 level.

Figure 3. Words Linkages Analyzed by Voyant Links

Note: Paris dataset; the linkages of the hotel and highly relevant words are highlighted in dark gray

Results by Co-occurrence Network

This tool by KH Coder is a software program providing several analysis tools for quantitative content analysis and text mining (Higuchi, 2015). KH Coder was originally developed for Japanese text and now supports numerous other languages, including English, Italian, French, Spanish, and even other Asian languages like Korean and simplified Chinese (KH Coder Index, 2012). Co-occurrence network provides a diagram of frequently occurring words in a text by identifying nodes and edges. The network diagram displays words and edges that indicate having similar appearance patterns to other words. Each node size indicates the frequency of words in a network diagram and thickness of edge specifies the presence of co-occurrence between words in text. However, the length of edges linked between words does not designate the degree of combined association strength. Word frequency list and statistics of collocations are provided as additional information and KH Coder provides a manual function to exclude stopwords.

Co-occurrence structure can be visualized according to degrees of centrality and communality as centrality-degree and communality-between. Centrality-degree indicates how central the word is in a network. It shows the keywords with similar appearance patterns particularly regarding centrality (lighter gray = low centrality; gray = high centrality). Communality-between indicates how words in the same community can be detected in a subgroup. Distinctive communities between words can be identified with different colors. In theory, results are similar to those generated by a cluster analysis (KH Coder, 2017).

For example, in the diagram for centrality for 5-star hotels in Hong Kong, while “room” is the most frequently mentioned keyword, followed by “hotel,” their degrees of centrality are shown in different shades of gray. While “room” is more frequently mentioned than “time,” the result shows that the most centered keyword is “time,” as shown by the gray color which is darker than the other keywords. Except for “time,” “room,” and “hotel” are at the next level of high centrality words, followed by “stay,” “staff,” and “day.” The other words in lighter gray specify the lower degree of centrality in the co-occurrence networks. Moreover, 10 different clusters are identified in the extent of gray color in a diagram for communality. The two results can be interpreted in two ways: the relationship between keywords by identifying the centralized words, and the clusters identified among the keywords. Co-occurrence Network Maps for “Centrality-Degree” and “Communality-Between” are presented in Figure 4.

Figure 4. Co-occurrence Network Maps

Note: Hong Kong dataset; the colors in the upper diagram indicate the degree of centrality and the actual colors range from gray to lighter gray in the outcome. Each separate group in the bottom diagram is highlighted in different colors based on communality between groups

Results by WordWanderer

This tool enables researchers to wander through words in the tools to create clouds in several visualization ways. Results in the interface are provided in three main ways (About WordWanderer, 2017). The first visualized presentation is a cloud view which is an overall picture displaying concordance. The most common words of a corpus are displayed to represent relative frequency according to different font size. The concordance is shown according to the alphabetical order from left top to right bottom. In hovering over a word, all related ones are highlighted in different color on the program, in terms of its rate of proximal co-occurrence in a corpus, while nonrelevant ones are shaded. The darker the words, the more frequently they were mentioned in the corpus.

In order to explore outcomes of WordWanderer from the Bangkok dataset, cloud, context, and comparison views are available in Figure 5. As with the other previously mentioned tools, “hotel” and “room” are shown to be the most frequently mentioned words. This tool is also useful when highlighting a particular word and exploring the relationship between two. For example, the two negative words (“bad” and “disappointing”) are selected. The results show all the common keywords associated with disappointing experiences in the reviews of 5-star hotels in Bangkok, namely,

Figure 5. Context View, Cloud View, and Comparison View

Note: Bangkok dataset; context View (upper left), All Associations with “Disappointing” in Cloud View (upper right), All Associations with “Disappointing” in Context View (bottom left), and All Combined Associations Between “Disappointing” and “Experience” in Comparison View (bottom right)

expensive, bathroom, check-in, people, staff, and taxi. The size of words corresponds to the strength between two words and horizontal position indicates differences in association between two. As a word occurs more frequently, they are placed more to the left side of the map. However, an overall relationship in a picture cannot be captured at once and only individual relationships are explored by clicking on one word at a time. In addition, the map cannot contain all words in a corpus, so only a limited number are included.

The Features of Five Visualization Tools

In order to identify the unique features of five data visualization analysis tools (Voyant Links, Voyant Cirrus, Lexos—BubbleViz, KH Coder-Co-occurrence network, and Wordwanderer), strengths and weaknesses of the five data visualization tools are explained in [Table 2](#).

In summary, each of the approaches can be distinctively applied in hospitality and tourism businesses because of at least five reasons. One, the result in a corpus of Voyant Links can be an informative but primary data outcome. It can be employed to visualize key relations highlighted with

Table 2. Strength and Weakness of Five Data Visualization Tools

Tools and Web Links	Strength	Weakness
Voyant Links http://voyant-tools.org/?view=CollocatesGraph	<p>(1) Diverse ways to create a corpus exist, such as text type or file type to upload (in either a single file or multiple files).</p> <p>(2) A collocation-network graph is hovering so that it is easy to manipulate the design and the shape of graph and any word can be placed at the center by dragging and dropping words according to the researcher's intention.</p> <p>(3) Unnecessary words can be removed in a graph by the user.</p> <p>(4) A highlighting function linking an association with a word is highlighted when the particular word is pointed out, so that it is easy to fetch collocating words.</p> <p>(5) Diverse visualized outcomes can be generated by adjusting the level of context ranging from 3 to 30.</p>	<p>(1) Each linkage does not include other advanced information instead of nodes and edges in a graph.</p> <p>(2) Results are displayed differently, depending on level of context in displaying a map of collocation.</p>

Table 2. (Continued)

Tools and Web Links	Strength	Weakness
	(6) Outcome table consists of further information in a form of summary table including total number of corpus, number of unique word forms, average number of words per sentence and most frequent words in the corpus as well as word occurrence frequency.	
Voyant Tools: Voyant Cirrus http://voyant-tools.org/	(1) A colorful picture of the corpus constituting a large amount of words can be generated easily and quickly. (2) Inclusive words can be decidable by adjusting the level ranged from 25–500. (3) All primary information is also identified in the default multi-tool skin.	(1) Words and colors in a word cloud may appear in different ways, in turn, the colors of words and the absolute positions of the words do provide significant meaning. (2) It is too simple to include key information.
Lexos: BubbleViz http://lexos.wheatoncollege.edu/upload	(1) Diverse ways to create a visual bubble map are available by downloading the program and	(1) Colors do not imply any particular meaning and overall shape of map also does not imply any precise meaning.

Table 2. (Continued)

Tools and Web Links	Strength	Weakness
	analyzing data on its server.	(2) Collocation between words cannot be identified from the relationship between bubbles.
	(2) Multiple documents can be saved in its database tool so that they can be switched from one document to another easily	(3) According to setting in analyzing data (e.g., graph size, maximum number of words, minimum word length), the bubble map provides different maps.
	(3) A very colorful and visualized picture (of bubbles) is presented.	
	(4) Graph size, maximum number of words, minimum word length can be adjusted to generate a bubble map.	
	(5) Four stages of data analysis are provided to scrub text by several default options, such as removing all punctuation, making the text lowercase and removing digits, including/excluding Keep Words/Stop Words.	
KH Coder: co-occurrence network http://khc.sourceforge.net/en/	(1) It particularly provides a manual function to exclude stop words.	(1) As KH Coder is a software, some functions cannot be securely operated although it is provided as a well-developed program.
	(2) Co-occurrence network graph can be generated according to degrees of	

Table 2. (Continued)

Tools and Web Links	Strength	Weakness
Wordwanderer http://wordwanderer.org/	<p>centrality and communality as Centrality-degree and Communality-between.</p> <p>(1) Outcomes are visualized in three main ways: cloud view, context view, comparison view.</p> <p>(2) Concordance and all examples of collocation of a word are provided.</p> <p>(3) Positions of words in each of the three views are shown according to certain orders for example, alphabetical order or sentence position related to the chosen word.</p> <p>(4) A comparison view is useful in highlighting a particular word and another word, and exploring relevant words and relationships between two words: for example, some negative words can be selected to see the relationship and the</p>	<p>(1) It provides a less simple and less colorful map.</p> <p>(2) The map cannot contain all words in a corpus because it only includes a limited number of word.</p>

Table 2. *(Continued)*

Tools and Web Links	Strength	Weakness
	relevant words between two words selected by the researcher. (5) Several options are available to transform initial words to proper forms of data before visualization analyses, such as transforming all uppercase words to lower case words, including certain type of word classes and excluding common words such as the, is, at, in, and of, and search function.	

keywords from vast amounts of big data by clicking only a few times on the user-friendly platform. It allows researchers and practitioners who have limited resources about their hotel properties to generate basic statistics and simplified relationships. It highlights key terms as a primary outcome before conducting a complex visualization analysis. Two, Voyant Cirrus can provide a simplified but visualized word cloud that is a useful figure and eye-catching diagram. The word cloud can analyze a huge amount of big data in a very brief time.

Three, the bubble map of Lexos–BubbleViz can be employed to dissect the data block into each single keyword and analyze competitors in a city by drawing a contrast among multiple hotel properties to view which terms are stronger for each hotel category and to think through the strategic advantages of reinforcing or deviating from these kinds of apparent received views. It can be used to provide diagnosis of key issues among the data.

Four, since the KH Coder can provide a co-occurrence network in two distinct ways, this tool can be employed to focus on linkages among big data by emphasizing communal groups and centrality. This tool has been adopted to identify the linkages in tourism academic research (Tussyadiah & Zach, 2017), but it can also be adopted by practitioners who need to understand various target markets and their wants and needs. Finally, Wordwanderer provides three different presentations about the visualization outcomes. The cloud, context, and comparison views can be used to understand customized results by highlighting each single word on the outcome page. The results from Wordwanderer are more than just a figure and the unique function of it is to enable decisionmakers in hotel properties to highlight negative keywords. It is the most practical visualization tool to highlight negative keywords as practitioners should understand unsatisfactory experiences of customers as critical issues of customer satisfaction.

CONCLUSION

Customer online reviews in hospitality and tourism have gained increasing attention both from researchers and practitioners (Schuckert, Liu, & Law, 2015). The main objective of the chapter is to present the results of five data visualization tools to understand guests' negative experiences at 5-star hotels in five city destinations. The results from the five visualization tools revealed that room and hotel are the two keywords that customers use to express their dissatisfaction. Service and staff are also cited frequently in the five city destinations. However, several unique keywords are found for each city. For example, "check" and "time" were mentioned in customers' negative reviews only for hotels in Singapore. Further, co-occurrence network maps revealed the different patterns of relationships among keywords in the five cities.

The study found that each tool has its own way of presenting visual outcomes and practitioners or researchers should recognize the strengths and weaknesses of them. Not all visualization may be applicable at all cases; users need to choose wisely which techniques to use and when. Although words cloud by Voyant Cirrus and Bubble maps by Bubble Biz are useful to identify the keywords, the results mainly show the frequency, rather than revealing any relationships among keywords. On the other hand, Voyant links, co-occurrence network, and WordWanderer allow researchers to explore the relationship among important concepts or terms.

While the technology and capacity to collect and store large data are growing rapidly, the ability to analyze these data volumes is increasing at a much slower pace. Access to big data is useless unless one can turn it into valuable information. That is, big data fails to contribute to tourism intelligence if it is not able to be digested and its insights appreciated by the intelligent user.

Visual data exploration helps the user to be directly involved in the data mining process (Keim, 2002). The task of big data visualization is to recognize interesting patterns and correlations. Although many information visualization techniques have been developed over the last decade to support the exploration of large datasets, they are not discussed in the academic literature. Current visualization tools reviewed in this chapter may be most suitable for initial big data exploration. Because these tools have relatively low performance in scalability and interactivity, it would be more useful if the visualization tools offered an interactive visualization function. Advanced visualization tools that can handle semi-structured and various types of unstructured data would be worthy of further exploration. Magnini, Honeycutt, and Hodge (2003) pointed out that various forecasting and research techniques should be used together with data mining, because the latter cannot capture all the information about what drives consumer behavior. Big data should serve the researchers' purposes, rather than dictate processes (Chudnow, 2001). Therefore, in order for data analysis using visualization tools to be effective, it is important to include the human element in the data exploration process and combine the flexibility, creativity, and general knowledge of human intervention with the enormous storage capacity and the computational power of today's computers.