Obtaining Insights from Museum Review

Yupawan Vannavanit

Faculty of Business Administration, Kasetsart University

Email: fbusywv@ku.ac.th

Abstract

The study aimed to obtain insights from museum reviews regarding topics that were discussed among over 300,000 reviewers between 2007 and 2020. Additionally, given the review rating, the study aimed to investigate reasons for those who gave low review ratings. We used secondary data of museum reviews scraping from TripAdvisor. The topic modeling technique of Latent Dirichlet Allocation (LDA) was used to systematically extract topics from text. The computer algorithm assigned each topic a probability distribution over vocabularies and clustered each topic in an unsupervised mechanism based on saliency and relevance. We then visualized our result in an Inter-topic Distance Map. According to our findings, there were seven broad topics that reviewers talked about including "Art", "Service", "Dining", "History", "Visitor", "Animal" and "Vehicle". To be more specific, around 40% of topic contribution involved topics of "service", "visitor" and "dinning". These topics' expectations could be managed and improved by sets of marketing strategy in terms of the marketing mix to satisfy customer needs and feedbacks. While the monetary value of museum experiences was a concern for those who gave a low rating. Hence, we recommended the use of a customized pricing strategy which is likely to commensurate with customer expectation as well as increase museum operation efficiency.

Keywords: Museum; Museum review; Topic models

Introduction

Starting from 3rd century B.C., museums have exhibited humanity importance including culture, history, politics, and religion. Although the purpose of museum could be varied from entity to entity, simply put, a museum is an organization that permanently or temporary presents artifact

objects and stories. Generally speaking, significant objects that are in accordance with museum's mission are collected and publicly displayed. According to Walhmer (2018), there were approximately 55,000 museums in 202 countries worldwide. Although, some of these are in large international cities around the globe, approximately 33,000 museums located in the United State of America. In 2019, top 3 most visited museums are the world largest art museum "Louvre", "National museum of China" and "Vatican museum" respectively. For Thailand, according to the statistic from Prince Maha Chakri Sirindhorn Anthropology Center, there were 1,580 museums listed as of January 2021. This number went up sharply over the last two decades, thanks to the expansion from temple and college museum. Thai museums categorized by museum type and internal management is shown in table 1 below.

Table 1: Operating museums in Thailand in 2020 by museum type

Museum type	Number of museums
non-profit	51
school	university	325
temple	395
community	107
government	235
national	43
business	63
private	166
local	146

Source: Prince Maha Chakri Sirindhorn Anthopology Center

From business and economic perspective, museums could be categorized as part of the arts, entertainment, and recreation sector, referring to the classification from the U.S. Bureau of Labor and statistic. While, in Thailand, it contributed to the tourism industry which has been a significant driver of Thai economy. According to, Surawattananon (2019), tourism contributed around 17% to Thai GDP in 2019 and roughly maintained up to 2.94 million employments nationwide. Despite monetary contribution from museums' revenue being small, Thai museums have played an important row in preserving Thai culture and history. It fosters deeper understanding of authentic culture and natural heritage as well as sharing enjoyment for Thai people and foreign visitors. For educational purpose,

it allows researchers and learners to investigate and educate themselves with tangible and intangible assets of society and nature.

For tourist attraction, feedbacks and comments from visitors are an important information in understanding their thought and experience. It can help museums in boosting their visitors' satisfaction and managing their resources effectively. Thanks to the advancement of technology where online reviews from museum visitors could be systematically collected and analyzed with affordable cost. The size and frequency of review data could timely captured visitors' thoughts and behavior. We aim to study museum reviews from the largest sites for travelers review in current online community "TripAdvisor". To obtain insights from a vast stream of text data, we apply state-of-the-art topic modeling technique and summarizing keywords to understand what reviewers were discussing about and what make them gave low review rating.

Objective

The study aims to obtain insights from over 300,000 museum reviews between 2007 and 2020. Two main objectives are:

1. To obtain topics that were discussed among museums' reviewers

2. To investigate the reason why reviewers gave low review rating

Scope of research

The research focuses on using secondary data of museum reviews from TripAdvisor. We alternatively use social listening method to collect data and apply state-of-the-art topic modeling technique on visitor's digital footprint to capture insights from this big data.

Literature reviews

A rapid development in technology and a reduction in cost associated with internet access via personal devices leads to an explosive growth of digital information. This large stream of data is routinely captured by marketers. Harnessing this data including reviews allows them to learn more about how consumers feel, behave and response to their products and services. Wedel and Kannan (2016) critically examined the implication of marketing analytics through structured and unstructured data including an online review and digital information. Their result identified three directions for

analytics methods which are marketing–mix spending, personalization and customers 'privacy and data security.

Among several data types online, a myriad of online reviews has been recognized as one of the most beneficial and useful information due to their granularity, velocity and potential to reflect reviewers' experiences and thoughts. There are many studies explained the importance and benefits of word of mouth from online review including Rosen et al. (2014) and Pauwels et al. (2009). Although, reviews could be gathered offline, Kane and Alavi (2014) revealed that a social network analysis from an online media has considerable impact than an offline network. Later, Schweidel and Moe (2014) suggested that sentiment of a media post could serve as a leading indicator of the changes observed from external data. They also recommended that focusing on various external data sources outperforms from single channel of social media post. Another key benefit of tracking online reviews is shown by Lamberton and Stephen (2016). They tracked changes of keyword counts for social media, digital and mobile marketing from 2000 to 2015 and suggested that marketing themes from these three channels could be used as a facilitator of individual expression, decision supporting tool and a market intelligence source .

Data driven marketing from online information potentially provides insights to corporates and companies that have capacity to leverage technology and reap benefits from this large amount of high frequency data. Rosario et al. (2016) and Berger (2015) showed that word of mouth affects consumer behavior and decision to purchase. Later Packard and Berger (2017) suggested that language used in word of mouth differently shape that impact. They studied more than 1,000 online reviews and highlighted the important role of language and endorsement style from word of mouth.

For non–profit including the unique type of museum which is temple, Andreasen (2012) suggested that non–profit and social marketing represent most complex and difficult context in marketing activities. However, there are some studies shown the implication of online review for museum, for example, Kumar et al. (2015) conducted a field implementation to increase revenue and attendance for aquarium. They applied data envelopment analysis to overcome four key challenges for non–profit organization including increasing revenue without increasing ticket price, increasing attendance without compromising visitor satisfaction, increasing the impact of media investments without spending more and attract customer with long–term value potential. Arnett et al. (2003) developed a salience model for non–profit marketing, involving enhancing individual relationship and social exchange.

Despite many benefits from analyzing customer reviews for both profit and non-profit organization being described above, the analysis should be done carefully. Anderson and Simester (2014) analyzed more than 12,000 reviews for retail firms and found that deceptive reviews not only limited strategic actions of firms but also lead individual customer without financial incentive to influence review ratings. Hence, the process of collecting, cleansing, and validating data should be done appropriately to avoid bias result. As there are limited number of literatures specifically discussed about online review regarding museum, the paper aims to explore this area of study by directly focusing on finding insights from museum reviews.

Regarding online museum marketing, it is important to increase website traffic with search engine optimization, applying geofencing marketing and monitor digital marketing according to WebFx (2021). Graham (2020) suggested that museum should encourage visitors to get social, allowing them to share and generate content online. She also mentioned that user-generated content is one of the most cost-effective ways to boost museum marketing. In addition, according to Manoukian (2019), museums should embrace technology without losing sight of their mission since the number of museums's attendance across the United States is falling. Museums are pillars of culture and learning where people can explore different traditions, new ideas, and unique art forms, according to Graham (2020), the paper aims to leverage museum reviews to find new insights and offer value to museums in terms of optimizing marketing strategy in data-rich environment. We aim to harness online data via TripAdvisor. This vast stream of digital information allows us to understand how reviewers feel, behave, interact, and respond to museum experience through their museum reviews.

Methodology

To obtain insights from museums' reviews, we first scrape data of TripAdvisor reviews from official TripAdvisor website. Our data contains 313,084 reviews from 2007 to 2020, representing museums experience across the globe. Data type and description are shown in table 2 below. We then explore the data set to detect anomalies and observe pattern. During this process, we remove "Stops words" but keep words related to overall museum context for the benefit of context understanding. We also lemmatize the data, cleansing them before performing two analyses namely, topic modeling and summarizing keywords by rating.

Table 2: Data description

Data	Data type	Data description
Museum	Character	Museum name
Rating	Character	Reviewer's rating ranging from 1–star to 5–star
Title	Character	Short description summarizing content of the review
Content	Character	Reviewer's review content

Starting with topic modeling, it is a generative model of language and text. Topic modeling is a type of computer algorithm that extract topics from text. As topics are automatically emerged from original text, they do not require prior annotation and document labelling. In terms of process, we first reshape our data into the form of a document–term matrix. Next, Latent Dirichlet Allocation (LDA), which is the baseline topic model, was used to cluster reviews with similar topic. From a collection of review, LDA assigns each topic a probability distribution over vocabularies and to each document a distribution over topics in a totally unsupervised mechanism based on saliency (Chuang et al., 2012) and relevance (Sievert & Shirley, 2014). The salient terms and relevance metric equation are shown below. After LDA topics' identification, we label those topics, according to their words' frequency and context relevance. In other words, our methodology is similar to machine learning version of focus groups.

$$\text{Saliency (term } w) = \text{frequency}(w) * [\text{sum\_t } p(t|w) * \log(p(t|w)/p(t))]$$

$$\text{Relevance (term } w \mid \text{topic } t) = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$$

t: topic

w: word

λ: the weight given to the probability of term under given topic.

Another analysis is to summarize keywords by rating to understand what drives reviewers' low rating. To do so, we first group reviews by the rating given by reviewers which represented by number of stars, ranging from 1–star to 5–star. We then remove "Stops words". Next, we rank top words, focusing on words which are noun for the purpose of interpretation. For each rating, we draw a conclusion based on words' frequency and context relevance.

Results

The result of LDA topic models is visualized in an Inter–topic Distance Map in figure 1. This interactive web–based map illustrates topics as distributions over words. It consists of two–dimensional plane in which its center is assigned by computer algorithm. Topics are presented as circles which are determined by the similarity between two probability distributions. Area or size of each circle presents overall prevalence for each topic. The inter–topic distances between topics are determined by multidimensional scaling. There are seven unlabeled inter–topic resulted from LDA, without any significant overlapping area of topics, confirming by visualization in figure 1. Each circle consists of individual terms that are most useful for interpreting the selected seven topics. We choose examples of some relevant terms from each topic according to their topic–specific probability and business context.

The results of topic labelling, examples of relevant terms as well as percentage of tokens are shown in table 3.

Figure 1: Inter–topic Distance Map (via multidimensional scaling)
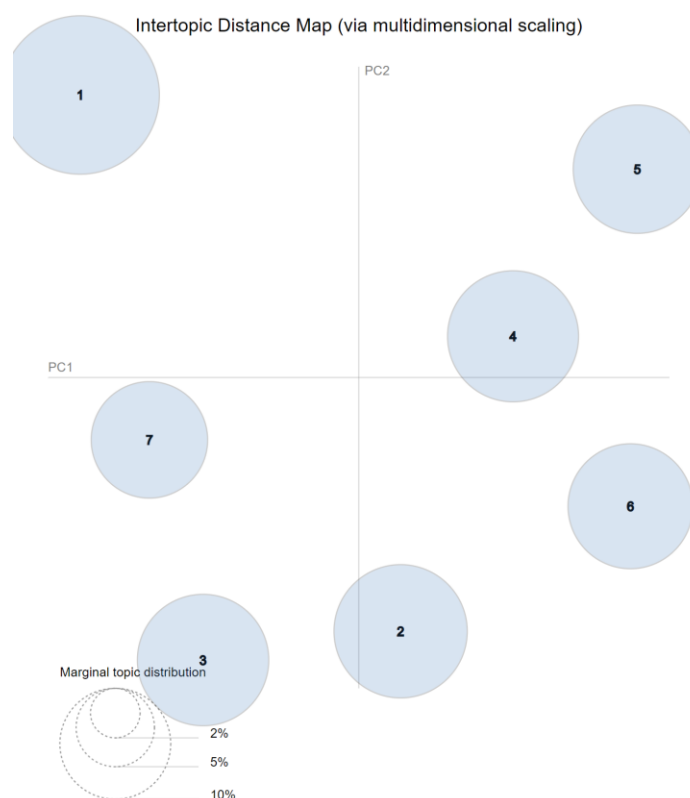
Table 3: Top words by topics of TripAdvisor museums' reviews

Inter–topic	Topic name	Percent of tokens	Examples of Relevant Terms for topic from most salient terms
1	art	20.4%	art, collection, painting, sculpture
2	service	14.5%	tour, guide, ticket, line
3	history	14.0%	history, life, time, culture
4	dining	14.0%	food, lunch, restaurant, cafe
5	visitor	13.4%	child, kid, family, adult, all–age
6	animal	12.7%	aquarium, fish, animal, turtle
7	vehicle	11.0%	car, bike, plane, motorcycle, truck

According to table 3, there are seven inter–topic, referring from figure 1, discussed by TripAdvisor museum reviewers. We label inter–topic from the context of relevant terms. "Art" is the most popular topic that were talked about among reviewers, accounting for 20.4% of total. Salient terms related to this topic includes "art", "collection", "painting" and "sculpture". Museum services and logistics were largely mentioned by reviewers capturing by words such as "tour", "guide", "ticket" and "line". This topic accounts for 14.5% of total tokens. Another important topic related to museum service is "dining", contributed 14% of total tokens. "Food", "lunch", "restaurant" and "cafe" are terms related to "dinning" topic. Topic related to "history" contributes as large as "dining" at 14% of total tokens. Some examples of relevant terms for this topic are "history", "life", "time" and "culture". The fifth most popular topic being discussed is "visitor" which consisted of words such as "child", "kid", "family", "adult" and "all–age". Another two important topics that reviewers were talking about are "animal" and "vehicle", accounting for 12.7% and 11.0% of total tokens, respectively.

The result of top words summarizing by rating, ordering by its frequency is presented in figure 2. Focusing on 1–Star and 2–Star rating, "money", "waste", "ticket" and "disappoint" are some words that stand out for these two rating. Beside high frequency words such as "museum" and "time", words such as "interactive" and "history" seem to stand out for high rating review.

Figure 2: Top words by TripAdvisor review rating
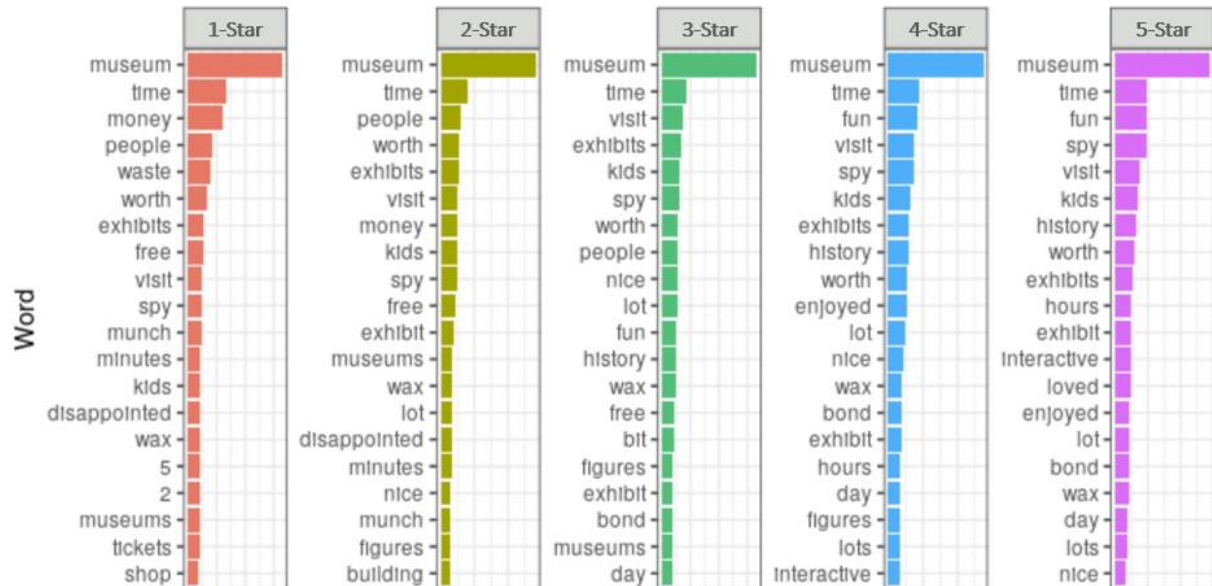
Discussion

Museums can be considered as unique destinations where most of museum visitors seems to have prior interest, and expectation about what they are paying for. These expectations could be different across individuals, countries, and museum types. Our result suggested seven distinct topics that were discussing among reviewers worldwide over the past decade. Museum type and artifact exhibited in the museum are important aspects that reviewer mentioned, combining contribution from these aspects including "Art", "History", "Animal" and "Vehicle" related topics accounts for 57.1% of total tokens. For practitioner, this part of contribution is mainly related to museum specific objective and goal. However, 42.9% contribution involves topics of "service", "visitor" and "dinning". These topics could be managed and improved by sets of marketing strategy in terms of marketing-mix to satisfy customer needs and feedbacks. For example, focusing on low-rating or low-star reviews, pricing and monetary value seem to be reason that make visitors unhappy. Hence, the improvement of pricing strategy to commensurate with customer expectation would both directly and indirectly impact museum operation. A direct impact is to improve visitors' museum experiences by fixing what they dislike, which is likely to result in higher rating overall. Regarding an indirect outcome, the spread of reviewers' positive word of mouth would influence others to revisit and encourage newcomers to visit, according to the finding from previous literatures.

Conclusion

"Art", "Service", "History", "Dining", "Visitor", "Animal" and "Vehicle" are seven topics that were talked about by TripAdvisor museum reviewers across the world between 2007 and 2020. These topics are well diverse for interpretation, where each topic approximately contributes between 10% and 20% of total tokens. In terms of keywords summarization by rating, monetary value seems to be a driver for 1-star rating as reviewers mentioned words such as "money", "waste", "ticket". As for 4-star and 5-star rating, distinct words such as "interactive" is stand out for those reviews.

Suggestion

Based on our finding, we recommend the use of various pricing options that offer visitors different museum experiences. Pricing should be customized with respect to visiting time horizon, artifacts accessibility, and the use of museum resources such as staff and space. Setting pricing alternatives should consider different museum expectations from various kinds of visitors such as tourists, artists, students, and families. For example, conservative spending visitors would prefer a campaign that allow them to pay for what they want to see with respect to the amount of time they spend. In contrast, "pay as you wish" pricing campaign would allow museums to benefit from charitable or kind visitors who are willing to give more.

Since museums usually do not allow physical contact with the artifacts, providing accessibility through virtual online exhibit with some interactive features will enable visitors to experience museums differently. Especially, during this pandemic where social distancing and strict lockdown measures applied nationwide, museums should leverage data and technology to advance visiting experience and alternatively allow visitors to see precious artifacts.

Based on our research methodology, we recommend the continuation of topic modeling method by apply it with museum specific data. This would allow a museum to insightfully understand their reviewers' thoughts and feedbacks which could significantly drive marketing decision and resource management.

New knowledge

Obtaining insights from online reviews by applying the state-of-the-art topic modeling allows us to understand reviewers' thoughts. This social listening technique widen marketer

perspective and deepen their understanding of reviewers' behavior and response. Broadly speaking, topics of museum specific including "Art", "Animal", "Vehicle", and "History" were discussed among reviewers, contributing around 57% of total tokens. Interestingly, 43% of museum service-related issues including "dining", "ticket", and "line" were mentioned as well. Taken together, museums should place an important on an improvement of guest service and necessary facility as they account for nearly 50% of what were discussed among reviewers. Another insight is that "pricing" is one of an important reason that, on average, make reviewers dislike museums.

References

Anderson, E. T., & Simester, D. I. (2014). Reviews without a Purchase. *Journal of Marketing Research*, *51*(3), 249-269.

Andreasen, A. R. (2012). Rethinking the Relationship between Social/Nonprofit Marketing and Commercial Marketing. *Journal of Public Policy & Marketing, 31*(2), 36-41.

Arnett D. B., German S. T., & Hunt S. D. (2003). The Identity Salience Model of Relationship Marketing Success: The Case of Nonprofit Marketing. *Journal of Marketing*, *67*(2), 89-105.

Berger, J. (2015). *Word of Mouth and Interpersonal Communication, the Cambridge Handbook of Consumer Psychology*. Cambridge, UK: University of Cambridge Press.

Chuang, J., Manning C. D., & Heer, J. (2012). Termite: Visualization Techniques for Assessing Textual Topic Models. *AVI '12: Proceedings of the International Working Conference on Advanced Visual Interfaces, May 2012, pp.74-77.*

Graham, A. (2020). *Attract More Guests with These Museum Marketing Tips.* Retrieved January, 15, 2021, from https://bigsea.co/attractions-and-museums/museum-marketing-tips

Manoukian, J. (2019). *Museum Marketing :Top Trends and Strategies for Success.* Retrieved February 10, 2021, from https://crowdriff.com/resources/blog/museum-marketing

Kane C. G., Alavi M., Labianca G., & Borgatti S. P. (2014). What is Different about Social Media Networks? A Framework and Research Agenda. *MIS Quarterly*, *38*(1), 275-304.

Kumar, V., Sharma A., Naveen, D., & Rountree C. (2015). Implementing Integrated Marketing Science Modeling at a Non-Profit Organization: Balancing Multiple Business Objectives at Georgia Aquarium. *Marketing Science, 34*(6), 804-814.

Lamberton, C., & Stephen A. T. (2016). A Thematic Exploration of Digital, Social Media, and Mobile Marketing: Research Evolution from 2000 to 2015 and an Agenda for Future Inquiry. *Journal of Marketing, 80*(6), 146–172.

Packard, G., & Berger J. (2017). How Language Shapes Word of Mouth's Impact. *Journal of Marketing Research, 54*(4), 572–588.

Pauwels, K., Trusov, M., & Bucklin R. E., (2009), Effects of Word–of–Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing,* 73, 90–102.

Prince Maha Chakri Sirindhorn Anthopology Center. (2020). *Thai Museums Statistics.* Retrieved February, 10, 2021, from https://db.sac.or.th/museum/statistic

Rosario, B., Sotgiu, F., Valck, K., & Bijmolt T. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta–Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research,* 53, 297–318.

Schweidel, D. A., & Moe, W. W. (2014). Listening in on Social Media. *Journal of Marketing Research, 51*(4), 387–402.

Sievert, C., & Shirley K. E. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceeding of the workshop on interactive learning, visualization, and interfaces, June 2014, Baltimore, Maryland, USA. Pp. 63–70.*

Surawattananon, N. (2019). *Tourism and its role in driving the Thai economy; Is a hero necessary? A real hero?.* Retrieved February 10, 2021, from https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_29Oct2019.aspx

Rosen, E., & Simonson, I. (2014). *Absolute Value: What Really Influences Customers in the Age of (Nearly) Perfect Information.* New York: Harper Business.

Walhmer, M. (2018). *How many museums in the world?.* Retrieved February 10, 2021, from https://www.museumplanner.org/how–many–museums–in–the–world

WebFx. (2021). *Creative Marketing Ideas to Attract Museum Visitors.* Retrieved February 10, 2021, from https//:www.webfx.com/industries/tourism–hospitality/museums/marketing–ideas.html

Wedel, M., & Kannan P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing, 80*(6), 97–121.