Big data application in tourism – the case of Hong Kong

Paul Leung¹, Jamie Lo²

¹Technological and Higher Education Institute of Hong Kong, Faculty of Management and Hospitality, 133 Shing Tai Road, Chai Wan, Hong Kong
²The Hong Kong Polytechnic University, Hong Kong Community College, 9 Hoi Ting Road, Yau Ma Tei, Kowloon, Hong Kong

Abstract: This paper attempted to review how effective tourism entities are utilizing data especially big data collected from social media for their decisions, and reveal some of the perceived barriers in practical utilizations. The study adapted case study as its methodological design and make use of Hong Kong as the subject for investigation. Data and method triangulation will be used to ensure validity and objectivity. Key informants from different stakeholders including data generating institutes, research agents and consultants, IT specialists and users including government and travel trade will be interviewed to solicit their view on the subject. While Hong Kong will be used as the subject of investigation, the researchers believe that many of the finding will be generic enough to apply to other locality and even across different disciplines.

Keywords: Big Data, Social Media, Knowledge Management, Culture

1. Introduction

Mayer-Schonberger and Cukier (2013) exclaimed that big data has revolutionized our live, work and mode of cognitive process [1]. In the dynamically changing modern world, big data has critical impact on the decisions (Koscielniak and Puto 2015) [2]. Big data was defined as a consolidation of data from various sources, which include operational records and more importantly the postings on social media. In 2017, there were 2.46bn social network users worldwide and Facebook alone has recorded 2.1 billion active monthly user in Q4 2017. Among which, 1.7 billion were active mobile monthly active users (Statista, 2018) [3]. Active users were defined as those which have logged in to Facebook during the last 30 days. 1.40 billion people, however, logged onto Facebook daily in December 2017, which represents a 14 percent increase year over year and Over 300 million photos were being posted onto Facebook on a daily basis (Zephoria, 2018) [4].

The desire to generate insights and knowledge from data has a long history. The enthusiasm on data-driven decision-making has been ever increasing over the past 35 years (Agrawal et al. 2012) [5]. In the 1970’s, it was quite common for management to relate the quality of decision to the availability of information (for example, Feltham 1968) [6]. Information was believed to be critical for management effectiveness. Incomplete information, on the contrary, will not merely undermine the quality of the decisions but also affect the process and effectiveness of group decision making (for example, Kim and Ahn 1999) [7]. Therefore, efforts have been focusing on the provision of adequate and quality information. Management information system and technologies related have become more a critical subject with its own cumulative tradition (Culnan
Mithas et al. (2011) argued that an organization’s information management capability was directly correlated with its performance in various functional areas [9]. Researches also indicated that there was direct correlation between the deployment of IT and a firm’s profitability (for example, Barua et al. 1995; Mithas et al. 2012) [10] & [11]. While evidence suggested that there was general consensus on the importance of information, the provision of quality information has not matched with the growing needs (Buchanan and O’Connell 2006) [12].

Over the years, senior management has been focusing on creating necessary conditions for developing IT infrastructure and information management capability in order to improve the provision of data. With the advancement of latest technologies, capabilities in capturing and analysis data, and data-mining facilitations has improved the quality of information and, in turn, believed to be able to improve decision quality (Raghunathan 1999) [13].

In the experiment conducted by Keller and Staelin (1987), nevertheless, indicated a new dimension of concern [14]. The growth in the volume of information has imposed new challenges to usage. Their study indicated that the effectiveness of decisions was directly correlated with information quality but negatively with information quantity. Information overload would be detrimental to the potential contribution of data to decisions (Jacob 1977; Corte-Real, Oliveira and Ruivo 2016) [15] & [16]. When decision makers were overloaded with information, they would become confused, less accurate and thus their performance curtailed. In this regard, the provision of information has to be managed for effectiveness and efficiency. A people-centred information management system would be necessary for the optimizing the results (Marchand, Kettinger and Rollins 2000) [17].

The development of more recent information technologies and the application of big data on both paid and non-paid platform, however, seems to amplify and exemplify the situation. The developing trend of big data, cloud technology, social media and social listening are just a few factors to further complicate the situation. The overloading of information might create new barriers to the acquisition of knowledge and inhibited effective decision marking rather than facilitating them. Although new approaches such as big data analytics applications and BDA value chain seem to offer some possible remedies, these remained areas for further investigations on the one hand while the mentality and behaviours of users should be approached on the other hand.

The researchers conducted an initial investigation in mid 2016 on the collection and usage of big data among tourism industry in Hong Kong. This small scale study reviled that the collection of data, the process of information and the deep learning for the formulation of knowledge were neither efficient nor effective. Most the data being collected were archived but rarely being utilized for creating competitive advantages as suggested by Morabito (2015) [18]. In other words, the hospitality and tourism industry has not been fully utilizing big data in enhancing their competitive advantages and management effectiveness. The aim of this research is to investigate the reasons behind this situation and will attempt to suggest alternative solutions for rectifying the situation.

While this study has capitalized Hong Kong as the subject for investigation, the findings were believed
to have a broader application to other locality in the understanding of barriers in big data marketing.

2. Literature Review

Big data is generally defined as a set of large volume of data, which could be used to generate meaningful insight for facilitating management decisions. It has rapidly developed into one of the most critical domain for management science (see Jin et al. 2015) [19].

Big data concerns about how agencies and corporations collect, store, and analyze large quantities of data, and generate meaningful insights for decision making (Gantz and Reinsel 2011) [20]. It was argued as an ideal raw material of our modern economy.

Big data is signified by its size and complexity. These data, with proper processing and analytics, can contribute to the knowledge of the market and enable high quality decisions. It helps to better understand customers and create competitive advantages (Peled 2013) [21]. This is particularly important to the tourism arena given the rapid changing of the market landscape and highly diversified source markets.

The 7-Vs, which included volume, variety, variability, velocity, veracity, visualization and value, were the identified as important dimensions of big data’s operation (DeVan 2016) [22]. These dimensions, according to Lee (2017), were by and large technological [23]. He added a new dimension, Decay, which was referred to as the declining value of data over time. Lee (2017) systematically discerned the development of big data into three critical era and finally identified a list of challenges for the development big data [23]. Like a critical portion of the existing literature, it was largely related to the development of big data.

WTTC (2014) published a report on big data and its application in the tourism industry [24]. In its conclusion, it concerted three critical arguments. First, big data cannot replace creativity. Innovativeness and creativity will still be the core for creating competitive edge. Second, big data, the logistics for handling vast volume of data and the use of which have to be executed with special care. Third, big data requires new skills set for its benefits to reach an optimal level. In other words, the mere collection of a set of data with critical volume will not guarantee market success. In the playing field levelled with information, skills, mentality and creativity are required to transform information into knowledge, to generate insights and to build competitive advantage. “Consumer indifference to Big Data’s benefits must be carefully negotiated; and the data required will in many people’s eyes, cross a boundary of acceptability. One bad experience will negate 1000s of subtly improved experiences.” New skills and mentalities are also required for the decision makers and their organization to take the most benefits of Big Data. The question, therefore, is concerned with what kind of training will the future require and how long will it take to install these characters.

While big data seems to be promising for business, the adoption of big data and the translation of data into meaningful input for decisions would encounter challenges and barriers. In its earlier stages of development, the challenges are largely with the technologies and logistics to collect and process data. Elragal and Klischewshi (2017), for example, argued that big data collection is by nature passive and lack of focus [25]. While critical research should be built around defined problems, big data investigations are subject to the so-called ‘streetlight’ effect. Researchers has a tendency “to study phenomena for which there exist plethora of
In the 5th annual Digital IQ Survey conducted by PwC, results indicated that 62% of the respondents believed big data can facilitate the development competitive advantages. However, 58% of respondents were anticipating major challenges in generating insights from data (PwC 2013) [26]. DeVan (2016) argued that the biggest challenge to big data could be the ingesting and digesting of data in such a volume and diversified structure of which [22]. Olavsrub (2013) suggested that the four major barriers are limited capability in data visualization, budget limitation on data collection and analysis, talent gap and insufficient system capacity [27]. In the 2016 version of the PwC survey (2017), result indicated that the shortage of talents (63%), outdated technologies (61%) and lack of integration of new and existing technologies and data (59%) remain the major barriers to big data adoption [28].

Burgelman (2014) also contended that there are still important barriers inhibiting the adoption of big data for managerial decisions [29]. He cited the results of various user surveys and came to a hypothesis that big data was not being used as effectively as it should be. The three most important barriers to bid data adoption include the fear of new technology, lacking a change of mindset to deal with data and data system; and the lack of support from the top executive level. Alharthi et al. (2017) also concerted that big data adoptions require to overcome barriers in technology, people, and organization. In order to overcome these barriers, they suggested that changes in technology infrastructure, focus on privacy, promotion of big data and analytic skills development, and the creation of a clear organizational vision are required [30].

On top of the aforesaid, Lee (2017) further indicated the barriers of data security, investment justification and the issue of privacy [23].

These latest literatures’ arguments are valid and appropriate. They, however, often neglected the political and cultural aspects of the big data domain. The ownership of data, the proper respect to given to facts, conflicts of private interest, data integrity and struggle of power have rarely been addressed. These new dimensions will constitute part of the focus of this study.

As a summary, the existing body of literature on big data by and large concerned on two critical aspects of the subject, namely the development of data (technical aspect) and the benefits of big data (utility aspects). The development of knowledge in related to the users perspectives and barriers to utilization are relatively scant. Within which, the political concerns about the development and dissemination of knowledge are rarely being touched upon.

3. Objectives

In response to the development trend in big data and the rapid advancement of information technology, the researchers is interested to investigate three issues related to the application of big data in hospitality and tourism industry. First, whether decision makers still embrace the impression that insufficiency of information affected their decision quality. Second, what are the barriers in more fully utilize big data for management purposes. Third, what are possible alternatives for overcoming the barriers?

The researchers hypothesized that the critical barriers for the application of data in the development
knowledge and the facilitation to decision making go beyond technical dimension and found its root in socio-political domains. By identifying the critical impediments, it is believed that a more effective and efficient model of knowledge management can be propelled to enhance management effectiveness. While the results of this pilot study might not be able to provide solutions, the research can serve as a stepping stone to a meaningful discourse.

4. Methodology

This study utilized multiple methods for the collection of data. Owing to its exploratory nature, the research paradigm is based on critical thinking and qualitative investigations. Within which, in-depth key informant interviews contributed the major body of data. The research tried to assess the topic from both supply and demand side of big data. Therefore, representatives from both orientations were invited to participate. They included official data provision authorities, research agencies, academia, travel trade, hotel operators, tourists, and journalists.

Hong Kong has been selected as the locality for investigation. Hong Kong was selected for convenience and the ready access to critical key informants. While Hong Kong has its uniqueness as a tourism destination, the situation in related to big data adoption and cultivation of a knowledge based culture were believed to be generic. Besides, the nature of this study was exploratory in nature and the selection of subject should not be regarded as an obstacle in analysis.

In order to generate certain points of agreement, Delphi method was utilized to create communication amongst respondents and to consolidate concerted views. The identity of the respondents were disguised in order to protect their personal interest and to avoid embarrassment.

The researchers have developed a pool of qualified respondents and they were stratified into clusters, such as data providers and users; particular industrial segment and level in the organization. The initial pool included sixty-eight persons. By eliminating overlapping calibres, 45 invitations were sent out and 33 interviews were conducted.

The mode of contact was a mixture of personal, phone and written communications. The researchers tried to initiate the conversation with each respondent through face-to-face communication. It is believed that the face-to-face interview can provide more depth and details on the discussions of the topic. Owing to the time constraint, however, subsequent communications have adopted written communications via emails and/or through phone conversations.

Interview field text was transcribed into research notes and further content analysis were conducted to identify themes and attributes. Among the various sectors that the researchers have pinpointed for investigation, the local transportation operators have not been secured for interview.

5. Findings and Discussions

There was a general consensus among the key informant that information is critical for decision making and big data would be useful for their business decisions. This confirmed the existing literature
regarding the potential value of big data to operations. However, the research observed a general lack of knowledge about big data and what are the potential utilities of such data. When asked about the need and usage of data, information and intelligence, most of the respondents cannot differentiate the difference and they are expecting quick answers to their question.

The first part of the interviews focused on the perceived importance of data attributes. Respondents, especially those from the hotel industry, airlines and attractions were particular concerned about the availability of usable data. Owing to the nature of their business, they require up-to-date information and forecasts for them to make decisions, which are critical to their yield management and stock control. The quality of information, according to the informants, shall include availability, usability, trustworthiness, completeness and standardization. Most of them would like to have access to the data set and analysis the data by their own on a timely basis. These were largely in line with the 7-Vs of Big Data as featured in the existing body of literature. Results, however, identify new dimensions such as usability and standardization, which is the user friendliness of the produced information.

For example, many respondents commented that they have difficulties in comprehending information presented in different formats and having different definitions. Standardized format would, therefore, enhance the usability. Likewise, some of the current information gateways were perceived as complicated. Sometimes, they have difficulties in getting access to the same reports that they have been using. The general frustration was observed.

Respondents also agreed on the timeliness of the data/information, which is, to a certain, a mirror aspect of Decay of big data by Lee (2017) [23]. They raised the issue that the lead time for the data/information to be published was sometimes too long for creating critical values for their business decisions. Some respondents used the arrival statistics published by the authority as an example, which was by 60 days behind the calendar month. Users would have difficulties in making quick market decisions without the up-to-date information.

Volume of data, which has been a major characteristic of big data, has not been identified as a core attribute. Respondents from the users’ side would prefer reports in standard formats with management implications as takeaways to a large data set for their own analysis. With the exception of respondents coming from large hotel groups and airlines, most of the respondents suggested that they would not have the time and manpower to use a large data base.

Many of the respondents contented that they did not have access to the information that they would regard as critical. Even when they asked for the information, there was no guarantee that they could get what they wanted. That was the reason why availability has been identified as one of the most important attributes. Some of the respondents suspected that the reluctant in providing more information had political considerations. In order to protect the sources of data and the position of stakeholders, information and insights were sometimes withheld.

The suppliers of data and intelligence, however, shared a different view. They argued that their provisions were planned and should be sufficient to answer most of the users’ needs. An obvious discrepancy
perceptions has been identified. This finding indicated the needs of communication between the supply and demand sides.

Furthermore, suppliers also expressed their concerns about over-supply of information. Some of them argued that it was a Pandora box. Once they attempt to satisfy the data needs, more will be demanded for and there will be implications on cost and manpower requirements. More critically, their operations will have to answer questions regarding the data. In other words, workload will be escalated disproportionately when the provision of data started to increase in volume. In order to manage their cost and operational issues, they emphasized the needs to manage the provision of information and expectations of the users. They would tend to provide summaries of information in standard format and encourage users to stick with their provisions.

Some of the respondents from the supply side also suggested that it was the preference of users to receive simplified management implications and short answers rather than diving into the data set. The result of the study, however, indicated that there were critical differences among data users even if they are coming from different departments of a single company. Once again, better communication would be helpful to narrow the gap between the supply and demand sides.

According to information providers, both in-house and external sources, users were rarely capable to ask for the crucial information. Users had the tendency to use those data, sources and format of data that they are familiar with and refuse to explore new alternatives. The researchers called this the data paralysis syndrome. When this argument were presented to the users, many of them also agreed with the position. A couple of the respondents from the users suggested that there should be a better mechanism for suppliers to discuss with the users, understand their needs and provide trainings for facilitating their uses of data and information.

Both sides, however, share the concerns related to difficulties in data management, shortage of qualified manpower, and cost. For example, many respondents conceded that they do not have in-house capability to use big data and the cost to acquire qualified manpower can hardly justify by the cost. These have been in line with the existing body of literature.

The researchers then attempt to approach the political dimension of data and knowledge management. According to the technical respondents including research and insight specialists, data users and information providers often held a strong position on data collection methods and the results to be generated. They were often looking for evidence to support their decisions rather than know the truth. In other words, the decisions were virtually made before the collection and processing of data.

Some of the data users also contended with this argument. They suggested that business decisions were results of a complicated political process. The intelligence from the market could shed some light about the situation but might not be able to satisfy the organizational needs. Particularly among the middle level managers, they have to take the preferences of and directions from the senior level as guiding principles. If the data or intelligence were arguing for a different direction, they would have to fight an uphill battle, which might not be to their best interest. As a result, data users would sometimes prefer to work with the data set and to find “the preferred” evidence. Follow-on question was then asked about their ability to handle a large set of
data in order to find the “preferred” information that they are looking for. Most of the respondents agreed that they have the issue of capacity. They rarely being equipped with a capable technical team for data analytics. Yet, the access to the data set would enable them to pick those findings that are preferred and avoid those that are not.

Some of the respondents from the supplier’s side suggested that they were defaulted to provide key takeaway summaries to their simplest form and to avoid details. If they provide the users with the data set, they will not be in the position to control the development of their interpretations but to bear the consequences.

Information and insight providers also reviled that there are occasions when they structured the data collection and critically engineer the outcome so as to influence the users’ decisions. Therefore, data and the corresponding insights were engineered political tools rather than truth knowledge.

Data is not apolitical (Peled 2013). Politics played a critical part and influenced who is getting what, when and to what level of details. It has been common that collected data is processed, simplified and deferred in their publication by purpose so as to manage the consequences. Sharing of data was sometimes controlled without a clear and justifiable reason. Raw data set are rarely make available and thus meta-analysis was made impossible to users.

There are respondents from the supply side suggesting that they have been asked by top management to massage the data, alternate the collection method and “decorate” the findings. They suggested that data integrity were compromised and users might base their decisions on falsified evidence.

It was also critical to highlight the concern of ownership over data and knowledge. Respondents suggested that providers of information in general had a strong procession of ownership over the data and information that they provide. While users were in favour of gaining access to more data, suppliers were quite reluctant to share their data set.

As mentioned earlier on, users were rarely equipped with the skills and the software to comprehend the data and to conduct meaningful analysis. Given their time and manpower concerns, they would have to rely on insights being provided by a third party or a consultant so that they could focus on their decisions. Some of the respondents also suggested that another beauty of using a planning unit or consultant was that their accountabilities could be shared.

A critical barrier which might be more relevant to the government bodies and government subvention institute was their funding allocation mechanism. According to the existing system in Hong Kong, Government would earmark a recurrent funding for those continuing activities such as research. The amount was allotted based on the actual needs and will sustain into the future. The amount, however, will not be subject to inflation adjustment. The capacity of the investment, in such a case, will be jeopardised along the timeline. For example, the Hong Kong Tourism has been collecting tourism related information from various sources including the arrival data from the Immigration Department of Hong Kong, hotel related data from the Hong Kong Hotel Association, and other tourist behaviours from primary data survey, the departure visitor survey. The survey has been soliciting input continuously for over 25 years. There are about 300,000 samples being solicited every year. The budget, however, has not been increased to reflect the general inflation.
The continuing collection, analysis and storage of data concerned became burden to the organization. In order to mitigate the inflating cost, sample size has been trimmed for more than once over the years.

6. Conclusion

As a conclusion of the paper, the researchers attempt to generate corresponding strategies and tactics to cope with the aforesaid issues. First of all, the big data landscape has changed owing to the general awareness of value of data and the improvements in technologies. There is, however, still a gap in what can big data bring to users and how they could handle the mass volume of information. Therefore, training and education are sincerely needed.

Second, the data, information and knowledge generated must be valid and appropriate. Political influences and private interests have to be minimized in order to enhance the quality and contributions of data.

Third, given the fact that the tourism industry is still featuring a large composition of smaller organization, which would have difficulties in handling big data, it is suggest to establish a central platform for data collection, processing, knowledge generation and dissemination. This central tourism intelligence body should provide a one stop shop for the industry to get services from raw data, technical support on information usage, market information and intelligence.

Owing to the advancement of technologies, it is possible to set up a platform where customized report could made possible on top of the simpler digest versions. The platform’s interfaces, however, have to be user friendly and trainings have to be provided.

It is also recommended to enable this platform to collect data from users. Such as sharing the bookings, consumers' patterns of consumption and the questions that they have asked would be valuable to the development of knowledge and the empowerment of the platform.

Reference


**Author Profile**

Paul Leung received the Ph. D. from Bournemouth University in 2004. He has served the Hong Kong Tourism Board as Director of Strategic Planning and Research and the Hongkong Disneyland as Director of Research. Currently, he is the programme leader for the BA (Hons) in Hotel Operations Management at the Technological and Higher Education Institute of Hong Kong.

Jamie Lo acquired her Bachelor of Arts in Hotel and Catering Management and Master of Science in Hotel and Tourism Management from The Hong Kong Polytechnic University. She is also a Certified Hospitality Educator (CHE) of the American Hotel & Lodging Educational Institute. She started her teaching career in the School of Hotel and Tourism Management, The Hong Kong Polytechnic University. Before joining the Hong Kong Community College, the Hong Kong Polytechnic University, she worked as a programme manager in another institute, where she was responsible for the planning and development of hospitality and tourism, wine studies and sommelier, and banqueting management programmes etc.