Tourism Management Perspectives 23 (2017) 75–86

Contents lists available at ScienceDirect

Tourism Management Perspectives

journal homepage: www.elsevier.com/locate/tmp

Application of business intelligence in the tourism industry: A case study of a local food festival in Thailand

Thanathorn Vajirakachorn [a,*], Jongsawas Chongwatpol [b]

[a] Department of Tourism Management, School of Business, University of the Thai Chamber of Commerce, 126/1 Vibhavadee-Rangsit Road, Dindaeng, Bangkok 10400, Thailand
[b] NIDA Business School, National Institute of Development Administration, 118 Seri Thai Road, Bangkapi, Bangkok 10240, Thailand

ARTICLE INFO

Article history:
Received 11 November 2016
Received in revised form 3 May 2017
Accepted 13 May 2017

Keywords:
Business intelligence
Analytics
Tourism
Festival
Revisit intention

ABSTRACT

Local festivals and special events are known to have a great impact on the local community, society, economy, and culture. Massive data about products tourists purchase, services they experience, destination choices they evaluate, and accommodations they select at the events can be captured, but the key question is how to translate such data into meaningful information so that event organizers understand the behavior of tourists in order to increase their satisfaction and boost revenues and profits. This study outlines a way to integrate a business intelligence framework to manage and turn data into insights for festival tourism. This framework combines the architecture of database management, business analytics, business performance management, and data visualization to guide the analyst in drawing knowledge from the visitor data. A case study from a local festival in Thailand is conducted to demonstrate the practical validity of the proposed business intelligence framework.

1. Introduction

Local festivals and special events are known to have a great impact on the local community, society, economy, and culture especially in creating new tourist segments and improving the identity and image of a destination (Getz, 2008; Jang & Feng, 2007). Visitors attend the events for different reasons. First-time visitors are more tourism and travel oriented, more active travel planners; repeat visitors, on the other hand, are more recreation and activity oriented and are more positive in post-trip evaluations than first-time visitors. Some enjoy visiting the destination for aesthetic reasons such as sentimentality, memory, and a sense of belonging; others visit for more utilitarian reasons such as better knowledge of a geographic area or for selected activities (Li, Cheng, Kim, & Petrick, 2008; Quintal & Polczynski, 2010). Conducting surveys to profile tourists visiting the festival, gain insights into potential event tourist segments, create new events for specific target segments, understand why tourists decide to revisit the event, or modify the marketing mix of existing events is fundamental in developing effective tourism management and marketing strategies (Getz, 2008; Jang & Feng, 2007; Li et al., 2008; Quintal & Polczynski, 2010).

As tourists travel around the world to visit local festival events, massive data about products they purchase, services they experience, destination choices they evaluate, or accommodations they select can be captured. The key question is how to translate such data into meaningful information so that tourism service providers understand the behavior of tourists to increase their satisfaction or boost revenues and profits. This question is crucial because such data is usually analyzed in traditional ways through descriptive statistics or conventional Excel-based regression analysis. Consequently, not only are potentially important factors neglected, the results produced by traditionally applied statistical surveys may not appropriately represent and recognize patterns or behaviors of tourists visiting destination sites.

Fortunately, existing information technology has the capability to handle and support vast troves of data and content. Business intelligence is one of the application areas of growing importance in supporting business decisions. The concepts of business intelligence open the door to opportunities to integrate platforms to handle complex, unstructured data from emerging data sources and emphasize the analytical process of turning the data into actionable strategies for better business decisions (Phillips-Wren, Iyer, Kulkarni, & Ariyachandra, 2015). Although previous studies have employed business intelligence in exploring and analyzing large amounts of complex data in many different areas such as marketing, manufacturing, and supply chain management (Chen, Chiang, & Storey, 2012; Turban, Sharda, & Delen, 2011), the implications of the business intelligence schema in the tourism industry are yet to be well developed and established. This study seeks to fill this gap and outlines a way to implement a business intelligence framework to manage and turn data into insights for festival tourism. A case study from a local food festival in Thailand has been conducted to explore the implications of business intelligence and business analytics in the tourism industry. Specifically, the framework helps not only in answering the research question "What are the most important factors influencing visitors'

* Corresponding author.
E-mail addresses: thanathorn_vaj@utcc.ac.th (T. Vajirakachorn), jongsawas.c@ics.nida.ac.th (J. Chongwatpol).

http://dx.doi.org/10.1016/j.tmp.2017.05.003

intention to revisit the festival site," but also in understanding the key attributes that impact visitors' satisfaction regarding the services they experience at the local festival event.

The rest of this study is organized as follows. After a brief literature review in Section 2, Section 3 presents the proposed business intelligence framework of this study. Section 4 explains the festival events and how the data used to evaluate the proposed framework are collected. Results and discussion are presented in Section 5, followed by managerial and practical contribution in Section 6, challenges and lessons learned in Section 7, and the conclusion in Section 8.

2. Literature review

Business intelligence (BI) and business analytics (BA) have drawn attention in both academic and business communities over the past decades. Organizations view both BI and BA in different ways, from tools, techniques, technologies, and systems to practices, methodologies, and applications that help enterprises make better and more timely decisions by analyzing critical business data (Chen et al., 2012). Business intelligence is an umbrella term that combines architectures, databases, analytical tools, methodologies, and applications to aid in decision-making processes (Turban et al., 2011). The architecture of business intelligence consists of four main components: data warehouse, business analytics, business performance management (BPM), and user interface (Turban et al., 2011). Database management and data warehousing are considered the foundation of BI, as they are concerned with how data is collected, organized, stored, extracted, and integrated so that end users can easily view or manipulate the data in a timely manner. Business performance management (BPM) focuses on monitoring, measuring, and comparing a variety of performance metrics defined as the core tenets of a business strategy (Chen et al., 2012; Turban et al., 2011). Business analytics refers to the broad use of data and quantitative analysis, usually grounded in data mining and statistical analysis, to develop new insights and understand business performance (Chen et al., 2012; Davenport, 2010). Gartner, the world's leading information technology research and advisory company, identifies four types of analytics capability which help enterprises move from traditional descriptive analytics (what happened?), to advanced diagnostic analytics (why did it happen?), predictive analytics (what will happen?), and prescriptive analytics (how can we make it happen?) (Rivera & Meulen, 2014). Data mining techniques such as decision trees, neural networks, support vector machines, and cluster analysis are adopted for data segmentation, predictive modeling, association analysis, clustering, and classification in various business applications (Chen et al., 2012). A user interface (UI), usually referred to as a dashboard or data visualization, allows bidirectional communication between the system and end users and provides a comprehensive view of corporate performance measures (Turban et al., 2011).

Business intelligence and analytics have been applied in many fields across all industries, from customer relationship management, behavioral profiling, healthcare, and genome analysis to supply chains (Davenport, 2006; Kusiak, 2006). Mayer-Schönberger and Cukier (2013) and Minelli, Chambers, and Dhiraj (2012) illustrate the applications of big data analytics to enabling competitive differentiation to discover and solve business problems. The key concept of big data and analytics is to use analytical techniques to describe, explore, and analyze large and complex datasets that require advanced data storage, management, and visualization technologies (Chen et al., 2012). Chase (2009) provides an overview of demand-driven concepts including forecasting methods and performance measures. With advanced analytics through predictive modeling, companies can not only analyze data for timely decision making but also uncover patterns in customer purchasing behavior and evaluate their marketing campaigns accordingly. Davenport (2006) presents the value of business analytics and how enterprises use analytics to build competitive strategies and extract maximum value from their business processes. Many data mining techniques have been applied to improve inventory policies. For instance, Dhond, Gupta, and Vadhavkar (2000) present two case studies where the neural network technique is used for inventory optimization. The results show that at the same level of customer demand, the total level of inventory is reduced by 50%. Wang (2007) outlines the application of data mining in other areas in advanced manufacturing such as process and quality control, optimization of manufacturing yield, assembly selection, material requirement planning, and preventive machine maintenance. Ciflikli and Kahya-Özyirmidokuz (2010) develop a data mining solution for enhancing carpet manufacturing productivity. They employ attribute relevance analysis, decision trees, and rule-based induction, and the results indicate that the isolated machine breakdowns have been detected in the production process, and the proposed decision tree model shows a 72% improvement in the accuracy ratio. Narasimhan, Swink, and Kim (2005) apply cluster analysis to aid in examining the relationships between manufacturing practices and plant performance including new product development, flexibility, efficiency, and market-based performance.

Numerous studies have addressed various aspects of business intelligence applications in the tourism industry. Law, Leung, and Buhalis (2009) provide a comprehensive review of articles published in tourism and hospitality research journals regarding the evolution of IT applications, which can be grouped into the three categories of consumers, technologies, and suppliers (Law et al., 2009). Pyo, Uysal, and Chang (2002) outline how the discovery of knowledge in databases using data mining techniques can be applied in tourist destination management. The study discusses various aspects of knowledge discovery, from operational issues, tools, and techniques to applications regarding customers, markets, products and services, destination promoters, and tourism professionals (Pyo et al., 2002). Fuchs, Abadzhiev, Svensson, Höpken, and Lexhagen (2013) propose a business intelligence approach focusing on online analytical processing (OLAP) to illustrate how knowledge creation, exchange, and application processes for the Swedish tourism destination can be improved. The proposed knowledge destination framework outlines the integration of customer-based data sources, data extraction processes, data warehousing, and knowledge generation through data mining approaches (Fuchs et al., 2013). Kim, Wei, and Ruys (2003) apply an artificial neural network model to segment the market of West Australian senior tourists. Based on their demographics, motivations, and concerns regarding domestic and international holiday travel, all West Australian seniors can be segmented into 4 groups: 1) active learner, 2) relaxed family body, 3) careful participant, and 4) elementary vacationer (Kim et al., 2003). Bloom (2005) shows that the deployment of neural network models can enhance market strategies, especially in segmenting the tourist market for understanding changing behavior among tourists. For example, the model helps identify international tourists in profitable tourist segments who do not visit the travel site but have a profile similar to the tourists frequent the site (Bloom, 2005). Cluster analysis has also been used to segment tourists into subgroups based on their motivation to visit travel sites. For instance, a study from Park and Yoon (2009) finds four distinct segments among 252 tourists in Korea: family togetherness seekers, passive tourists, want-it-all seekers, and learning and excitement seekers (Park & Yoon, 2009). Lin and Huang (2009) employ a K-Mean data mining method to evaluate insightful patterns of destination images tourists consider when selecting a destination and to segment tourists' features into subclasses in the tourism market so that direct promotional campaigns toward specific classes can be launched (Lin & Huang, 2009). Kuo, Akbaria, and Subroto (2012) apply a particle swarm optimization algorithm to cluster Taiwanese tourists based on their motivation (such as cultural norms and values, family and reference groups, financial status, personality, and lifestyle) to visit Indonesia. The study also presents the preferred tourism destinations chosen by Taiwanese tourists, which range from heritage and culture to nature-based destinations (Kuo et al., 2012). Byrd and Gustke (2011) apply decision tree models to identify tourism stakeholders participating in tourism and political activities.

Research on identifying critical attributes impacting the experience of tourists at special events has been ongoing for decades. Mikulić, Paunović, and Prebežac (2012) combine a neural network model and work-based importance-performance analysis (IPA) to identify the attributes of visitors and exhibitors attending a regional wine fair in the town of Makarska, Croatia. The results of the study show that visitors perceive professionalism and fair attractiveness as the most important attributes of a positive fair experience; in addition, the size of the fair and the choice of the destination significantly influence the experience of visitors (Mikulić et al., 2012). Byrd and Gustke (2007) exploit a decision tree model to explore stakeholder involvement in tourism planning, development, and management by focusing on the perception of the impact of tourism on their community (Byrd & Gustke, 2007). Golmohammadi, Shams Ghareneh, Keramati, and Jahandideh (2011) propose a rough set-based neural network to predict tourists' overall satisfaction with their travel experience in Iran and to prioritize the importance of attributes that have a great impact on their satisfaction, which consequently influences the choice of destination and the decision to return (Golmohammadi et al., 2011).

These studies offer just a few of the implications of implementing the concepts of business intelligence and business analytics in the tourism industry. The business intelligence framework and the proposed 9 steps to accomplish this BI project in festival tourism are outlined in the next section.

3. Proposed business intelligence framework

A business intelligence (BI) framework to manage and translate data into meaningful information in festival tourism is proposed in Fig. 1. This framework combines the architecture of database management, business analytics, business performance management, and data visualization to guide the analyst in drawing knowledge from the visitor data

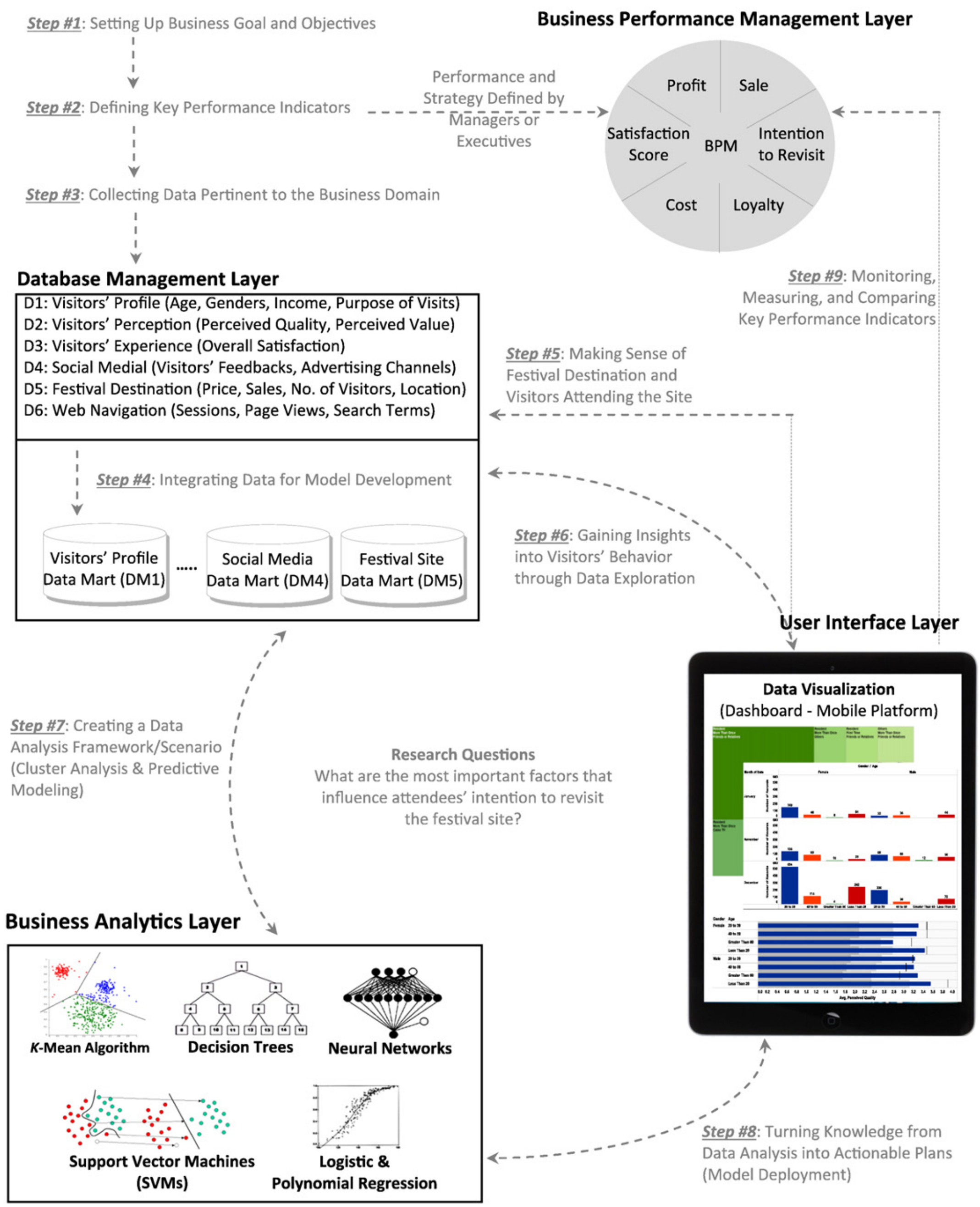


Fig. 1. Business intelligence framework.

(Rivera & Meulen, 2014; SAS, 2012; Turban et al., 2011). The first step is to set up the goals and objectives of the BI project in the domain of festival tourism. For instance, marketing analysts at the festival sites might be interested in developing predictive models to support managerial decision making with a focus on either understanding visitors' overall satisfaction with their travel experiences or segmenting visitors based on different attributes. The second step is to define the key performance metrics that are linked to the organizational strategies in the business performance management (BPM) layer. These indicators may include revenue, sales, profits, costs, satisfaction scores, or intention to revisit the festival site. The third step is to collect data that is relevant to the domain of festival tourism. For instance, in order to understand how visitors select the destination, it is important to include demographic-related attributes (age, gender, income, and marital status), motivation-related attributes (rest and relaxation, health, activities, or experience), and concern-related attributes (safety, language, culture, or hygiene) in the business analytics process. Step #4 in the database management layer is to integrate data that might come from different sources. This step usually takes time; it requires cleaning, formatting, and organizing data to prepare it for further model building, especially when the analysts face problems of data inconsistency, data duplication, and data errors. In Step #5, the data visualization layer is developed. It is important for the analysts to get some sense of how the festival destination is processed and to understand visitors attending the site. To aid in such data exploration activities, a dashboard, developed either on the web or mobile platforms, helps analysts first figure out what is happening to the key indicators defined in Step #2 and then understand visitor- and festival-related attributes though descriptive and frequency statistics. In Step #6, analysts look for insights into the data, whether they can detect any unusual patterns based on the preliminary data analysis. For instance, the analysts may find that approximately 50% of visitors attending the site are from the northern area of the country and only 10% of them received the advertised festival information through Facebook; on the other hand, only 20% of visitors are from the southern area of the country but over 80% of them found out about the festival site from the Facebook fan-page.

The business analytics layer, Step #7, focuses on creating data analysis scenarios where a variety of data mining techniques, such as cluster analysis, logistic and polynomial regressions, decision trees, and neural network models, can be employed to predict the targeted key performance indicators. These high-performance analytics through data mining approaches are applied and integrated to guide the analyst in drawing information from the data. Specifically, both data visualization in Steps #5 and #6 and business analytics in Step #7 help in answering the following types of business analytics questions: What happened, why did it happen, what will happen, and how can we make it happen. It is also important to set up a specific research question that is relevant to the business objectives defined in Step #1. For instance, the research question might be “What are the most important factors that influence visitors' intention to revisit the festival site?” or “What are the key attributes that impact the visitors' overall satisfaction regarding their experience at the site?” In Step #8, any information and key findings from the data exploration and analysis in Steps #5 and #6 and the predictive modeling in Step #7 that can be of value are used to provide feedback to the analysts so that any follow-up marketing campaigns to retain visitors for future events or actionable plans to improve visitors' satisfaction can be promoted. The last step, Step #9, is to monitor those strategies to ensure that the targeted performance indicators can be managed and controlled.

4. Methodology

The study was conducted at a downtown festival in Pattaya, Thailand, called “The 5th Walk to Remembrance at Naklua Market.” Naklua is an old fishing village located on Pattaya-Naklua Street, The event organizers started the festival in order to promote Naklua's local food and historical culture. The festival is held annually on eight weekends in November, December, and January, which is during the holiday season including winter break from school and New Year's celebration. Visitors enjoy the local culture, entertainment, food, and handcrafts as they walk through the historic market area with wooden houses along both sides. The estimated average number of visitors at the festival is approximately 1000 visitors per evening.

The dataset is derived from an estimated 3600 questionnaires collected from festival attendees (150 questionnaires per night, 3 nights a week for 8 weeks). The samples are selected by a stratified sampling method to avoid a site-specific bias based on the location of local visitors. The questionnaires were distributed to the festival sites, and respondents freely participated in answering the survey questionnaire after they had visited the events. The response rate was 56.8% (2048); however, of those 2048, only 317 questionnaires (15.47%) were complete and useable, without any errors (data duplication and extreme values). In addition to their demographic information, respondents were asked to rate their perceptions of the festival on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.” A total of 30 attributes regarding the demographic information, push factors, pull factors, perceived values, perceived quality, and satisfaction are used to explain and predict the intention to revisit the festival sites.

5. Results from applying the framework and discussion

5.1. Step #1: setting up business goal and objectives

Festival organizers are looking for ways not only to satisfy festival visitors so that they become repeat visitors, but also to gain new visitors based on advertising, promotion, or positive word-of-mouth from current visitors to their family and friends. Thus, the first task in the business intelligence (BI) framework is to set up the objective of this BI project: to uncover important factors influencing the intention to revisit the festival site.

5.2. Step #2: defining key performance indicators

With the increasing number of festivals held annually, it is critical for festival organizers to understand the behaviors of visitors. Visitors attend the events for different reasons and with different expectations; thus identifying their motivation to attend the event site, their perceptions of the value of the products they purchase or the quality of services they experience, and their level of satisfaction can help in assessing their revisit intention in the context of festival tourism. In this study, therefore, we want to develop mathematical models to determine which aspects of festivals and visitors' perceptions have the most impact on their intention to return.

5.3. Step #3: collecting data pertinent to the business domain

The data was collected from the “The 5th Walk to Remembrance at Naklua Market” event in November, December, and January of 2014. Approximately, 24,000 visitors attended the event, with the numbers climaxing on the opening and closing days of the festival. The organizers are able to explore the demographic information of the registered visitors: age, gender, and income. However, to leverage the positive impact of the festival event, the critical success factors that impact the intention to revisit the festival are measured based on how well the events meet the expectations of attendees. The important task for the event organizer is to identify the most critical festival attributes and provide the management with guidance to enhance the experience of future festival attendees.

The questionnaires collected information based on factors that had been found important in other literature. Many studies have focused on the antecedents of destination revisit intention as travel destinations rely heavily on repeat visitors. Quintal and Polczynski (2010) examine how perceived attractiveness, quality, value, and low risk impact university students' satisfaction and, consequently, result in their intentions to revisit a holiday destination. Um, Chon, and Ro (2006) outline some independent variables that contribute to revisit likelihood: attributes related to perceived quality of performance while on site, perceived attractiveness, perceived quality of service, perceived value for money, and satisfaction. Their results show that rather than overall satisfaction, perceived attractiveness is the most important indicator (Um et al., 2006). Jang and Feng (2007) explore the effects of tourists' novelty seeking and destination satisfaction to better understand temporal destination revisit intention (TDRI) measured on short-term, mid-term, and long-term bases. The results show that satisfaction is a direct antecedent of only short-term revisit intention and not mid-term and long-term revisit intention. Novelty seeking, a tendency to seek new and adventurous experiences, is a direct antecedent of mid-term revisit intention, which is connected to long-term revisit intention (Jang & Feng, 2007). Some might argue that for first-timers, intention to revisit a festival site is influenced mainly by the destination performance, while for repeaters, the most important factor driving their intention to revisit the festival site may be the promotional efforts to recall their positive memories or announce new attractions the festival is promoting. Although the determinants of revisit intention have been extensively studied, understanding tourists' revisit intention and their behavior remains an on-going research focus since most studies focus on different dimensions, with fragmented and sometimes conflicting findings that present a mixed picture of the critical festival attributes that impact the revisit intention. Additionally, each event has different characteristics, and the appeal of these events is hardly the same; tourists need to attend the events to differentiate and to enjoy the unique experience. Since the purpose of this BI project is to study visitors' behavior at the festival events, more data regarding perceived value, perceived quality, and satisfaction are needed. These variables are used to explain and predict the intention to revisit the festival in the future.

5.4. Step #4: data integration for model development

Since the data are obtained from different sources such as registration forms and questionnaires, the important tasks in this step are to organize the data into the same format, merge the data for model development in Step #5, and exclude any errors, inconsistencies, or outliers (extreme values) in the data. One of the biggest challenges in data integration is to handle missing values. Since not all visitors agree to participate in the study, the analysts encounter many incomplete responses. In preparing the data for analysis, only completed responses are used to avoid any missing information. After deleting responses with missing information, the dataset contains 317 complete observations for developing predictive models and cluster analysis to accomplish the goal of this study: identifying factors influencing the revisit intention of visitors. Additionally, because of the number of independent variables, multicollinearity is treated appropriately based on correlation analysis and variance inflation factors.

5.5. Step #5: making sense of festival destination and visitors attending the site

Exploring the data before model development enables the marketing analyst to get some sense of how the festival destination is processed or to understand visitors attending the site. To aid in such data exploration activities, a dashboard is developed either on the web or a mobile platform, the Data Visualization layer, using Tableau Software. The dashboard helps analysts first figure out what is happening to the key indicators defined in Step #2 and then understand visitor- and festival-related attributes through descriptive and frequency statistics. For instance, Fig. 2 presents the frequency distribution of female and male visitors based on age groups on November 30, 2013, December 14, 2013, and January 05, 2014, when the on-stage shows and the festival activities were similarly focusing on the culture and history of the

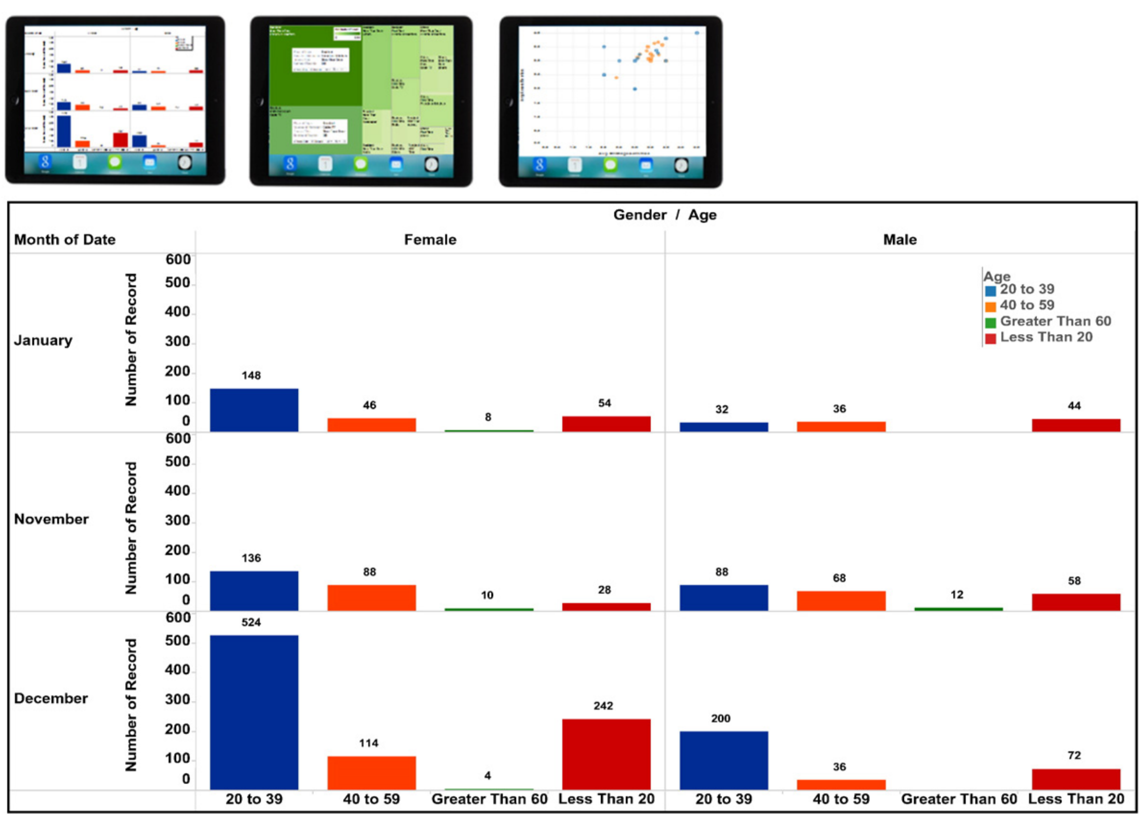


Fig. 2. Frequency distribution.

local community. Below are the key findings, which can be useful for the festival organizers in evaluating their marketing strategies and event preparation.

- More females than males attended the event on those three days: approximately 68% of visitors are female and 32% of visitors are male.
- The program on December 14 seems to attract more visitors.
- Most visitors (55%) are in the age group of 20 to 39 years old.
- Only 2% of visitors are over 60 years old.

If the purpose of this event is to engage all tourists from both inside and outside the community, clearly the organizers failed to draw the attention of senior visitors. Post-hoc analysis indicates that the content of the on-stage and activity programs are not attractive and the festival is not friendly for older visitors because of the inconvenient facilities such as restrooms, rest areas, and the location of the events, which is quite far from the parking lots and requires a long walk among the booths.

Fig. 3 also shows that most visitors (Points #1 and #2) on those three days have visited the festival more than once, reside in the community, and heard about the festival from friends, relatives, and cable TV. First-time visitors from other provinces (Point #3) attended the events based on friends' and family's persuasion. This finding indicates that the organizers need to reevaluate their advertising channels, since the current media—newspaper, radio, cable TV, website, and social media—may not efficiently and effectively approach new visitors.

5.6. Step #6: gaining insights into visitors' behavior through data exploration

The bullet graph in Fig. 4 is designed to compare the average perceived value (the black line) and perceived quality (the blue bar) of products and services both female and male visitors experience during their visit. Additionally, the scatter plot shows that first-time male visitors below 20 years old perceived the on-stage activities and general services differently from the female visitors over 60 years old.

5.7. Step #7: creating a data analysis framework/scenario

5.7.1. Scenario #1: cluster analysis

Cluster analysis is performed to get some sense of what visitors want to accomplish when they visit the festival site. The visitors are segmented into sub-clusters that share similar characteristics. Tourists decide to participate in the event for different reasons: One may just a rest or want to explore the festival with family and friend, another might want to learn about a new culture or meet new people at the festival site, and another may visit the festival because of the onstage activities and the reasonable prices of food and beverages. Thus, a k-Means algorithm is deployed in this clustering process in order to understand attendees' characteristics before building predictive models. The k-Means algorithm has been widely used in market segmentation because of its ability to identify distinct clusters in a dataset with multiple variables (Jackson, 2002; SAS, 2012; Turban et al., 2011).

5.7.2. Scenario #2: predictive modeling

Four popular classification platforms (neural networks, stepwise logistic regression, decision trees, and support vector machines) are used in this study to analyze the dataset with its multiple predictor variables. The target variable (Y) is "Intention to Revisit"; the independent variables include the following aspects:

- Demographic information: Gender (X_1), Age (X_2), Place of Origin (X_3), No. of Visits (X_4), and Sources of Information (X_5)
- Perceived value: Expectation (X_6), Value (X_7), Products & Services (X_8), and Price (X_9)
- Perceived quality of event: Location (X_{10}), Duration (X_{11}), General Services (X_{12}), Sellers (X_{13}) Onstage Activities (X_{14}), Traffic Management (X_{15}), and Facilities (X_{16})
- Pull factors: Variety (X_{17}), Reputation (X_{18}), Promotion (X_{19}), Accessibility (X_{20}), and Overall Experience (X_{21})
- Push factors: Learn New Culture (X_{22}), Meet New People (X_{23}), Find New Experience (X_{24}), Travel with Family (X_{25}), and Take Some Rest (X_{26})
- Satisfaction: Impressiveness (X_{27}), Right Decision (X_{28}), Quality of Activities (X_{29}), and Overall Satisfaction (X_{30})

Fig. 3. Visitors' profile.

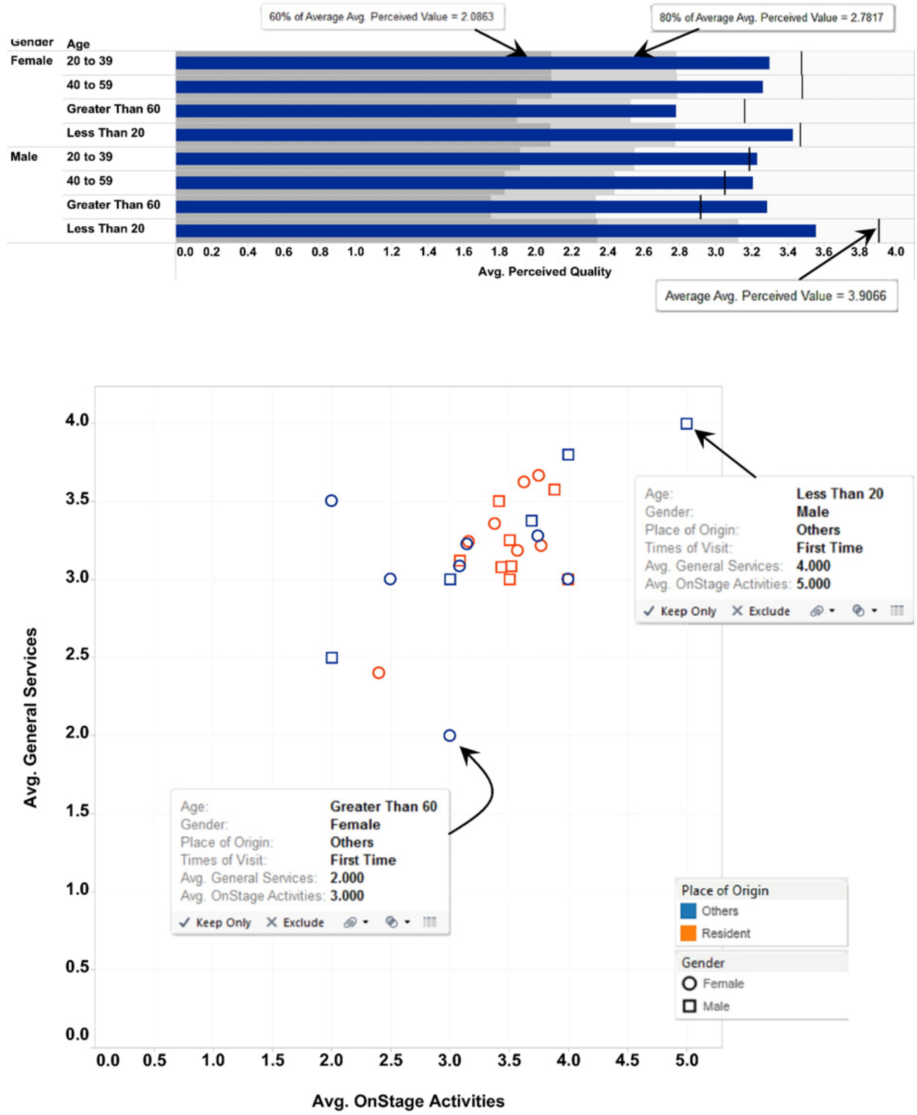


Fig. 4. Scatter plots and bullet graph on perceived quality.

Each of these four classification techniques is developed to build models to predict the desired binary target outcomes when "Y = 1" is Yes (Plan to revisit the festival event in the future) and "Y = 0" is the opposite (Do not plan to revisit…). Some of the techniques require a specific form of data and are based on certain assumptions. These techniques can be applied to the same problem types and are explained as follows:

An artificial neural network (ANN), which is also called a circuit of biological neurons, refers to a mathematical and computational model used to recognize or classify pattern in the data through a learning process. A biologically inspired analytical technique, an ANN model simulates a biological system or a brain's nervous system, where a learning machine algorithm indicates how learning takes place and involves adjustments to the synaptic connections between neurons. ANNs have been used successfully in many fields because of their remarkable ability to derive meaningful pattern recognition or data classification from complicated or imprecise data. Data input or independent variables can be discrete or real-valued; the output or targeted variable is in the form of a vector of values, which can be either discrete or real-valued as well.

Stepwise logistic regression is often used to predict a binary targeted variable or multi-class dependent variables with an automatic selection of predictor variables. It builds a model to predict the odds of discrete variables (dependent variables) using a mix of continuous and discrete predictors instead of point estimate events as in traditional linear regression models because the relationship between dependent variables and independent variables is non-linear. The emphasis in the stepwise procedure is on selecting predictor variables that best explain a particular predicted variable on the basis of statistical criteria. Stepwise polynomial logistic regression, an enhanced regression process, is used to predict a dependent variable on the basis of "n" independent variables. It is commonly used when the relationship between target variables and the explanatory variables is a complicated non-linear phenomenon.

Support vector machines (SVMs) are among the best supervised learning algorithms that are based on the concept of decision planes defining decision boundaries. An SVM is a classifier method that uses the mapping function to construct hyperplanes in a multidimensional space to either categorize the data for the classification tasks or to estimate the numerical value of the desired output in the regression tasks. Various kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid are used to transform the input data into the higher-dimensional feature space in which the algorithms output an optimal hyperplane so that input data becomes more separable. A larger minimum distance in the hyperplane indicates less error in classifying the input data. For the classification tasks, for instance, new observations are mapped into the same space and are predicted into the side of the gap they belong to. An SVM can handle several thousand training examples.

A decision tree is a well-known data classification and prediction method commonly used because of its intuitive explainability characteristics. A decision tree divides the dataset into multiple groups by evaluating individual data records according to their attributes. In other words, a decision tree provides the unique capability of splitting a dataset into branch-like segments with a root node at the top of the tree. It is also a simple and easy way to visualize the process of classification where the predicates return discrete values and can be explained by a series of nested or rule-based if-then-else statements.

For a technical summary, including both algorithms and their applications for each method, see (Harding, Kusiak, & Shahbaz, 2006; Jackson, 2002; Turban et al., 2011).

5.8. Step #8: turning knowledge from data analysis into actionable plans

5.8.1. Scenario #1: cluster analysis

Visitors to the festival can be divided into four groups based on the k-Means clustering analysis. Fig. 5 illustrates the three most influential variables used in classifying visitors into the clusters, each of which has a preferred reason for visiting the festival site. For instance, Cluster #1 comprises visitors who favor the variety of products, the prices, and the timing of the festival. Visitors in Cluster #2 are mostly satisfied with the traffic management of the festival organizer, the value, and the facilities at the site. Visitors in Cluster #3 emphasize the value and quality of products offered and attend the festival to have new experiences. Visitors in Cluster #4 tend to have high expectations of the festival and are mostly satisfied with the value and the quality of the festival activities. These four clusters provide an overview of different profiles so that the event organizers understand visitors' motivation and their purpose in visiting the site.

Additionally, further segmentation the profiles provides valuable insights into visitors' preferences or perceptions of the festival. For instance, Fig. 5 shows a multi-dimensional diagram for evaluating the variables particularly related to push factors for all four clusters. Event organizers can approach the visitors in Cluster #4 by promoting the value and activities the festival offers rather than emphasizing the experience and culture visitors can learn from. Attending the festival, for visitors in Clusters #1 and #2, increases the opportunity for spending time with their family; thus event organizers may come up with campaigns emphasizing that children can gain new experiences and learn about culture or promoting the easy access to festival site facilities (rest areas or restrooms) that are suitable for both adults and children.

5.8.2. Scenario #2: predictive modeling

5.8.2.1. Neural network model. Fig. 6 presents the pictorial representation of the neural network architecture developed to explain and predict the visitors' intention to revisit the festival site. This neural network model consists of one hidden layer and three hidden units with various input variables. The results of the prediction model are presented to show that each hidden unit (H_1 to H_3) has its own function and weights. The transfer function, also called the tanh function, combines all the inputs into a single value for each hidden unit and then calculates the output value of Intention to Revisit (Y). The following seven variables have high coefficient estimates in the neural network model and are among the important variables influencing the decision to return to the festival: Accessibility (H_{20}), Onstage Activities (H_{14}), Overall Satisfaction (H_{30}), Quality of Product (H_8), Age (40 to 59) (H_2), Travel with Family (H_{25}), and Traffic Management (H_{15}).

5.8.2.2. Decision tree model. The following four rule-based "if-then" algorithms illustrate the implications of the decision tree model.

- *IF* "Overall Satisfaction" is greater than 3.5 *AND* "Accessibility" is greater than 3.5 *AND* "Place of Origin" is resident, *THEN* the probability of returning to the festival is 76.9%
- *IF* "Overall Satisfaction" is greater than 3.5 *AND* "Accessibility" is less than 3.5 *AND* "Place of Origin" is not resident *THEN* the probability of returning to the festival is 37.5%
- *IF* "Overall Satisfaction" is less than 3.5 *AND* "Reputation" is greater

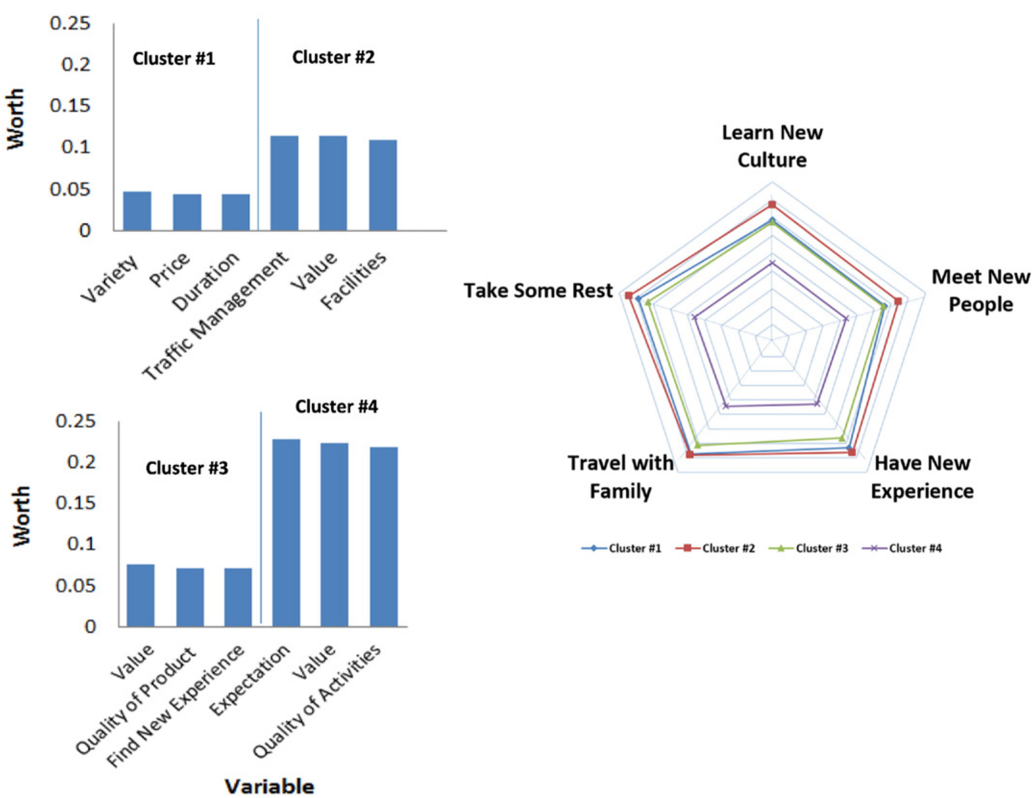
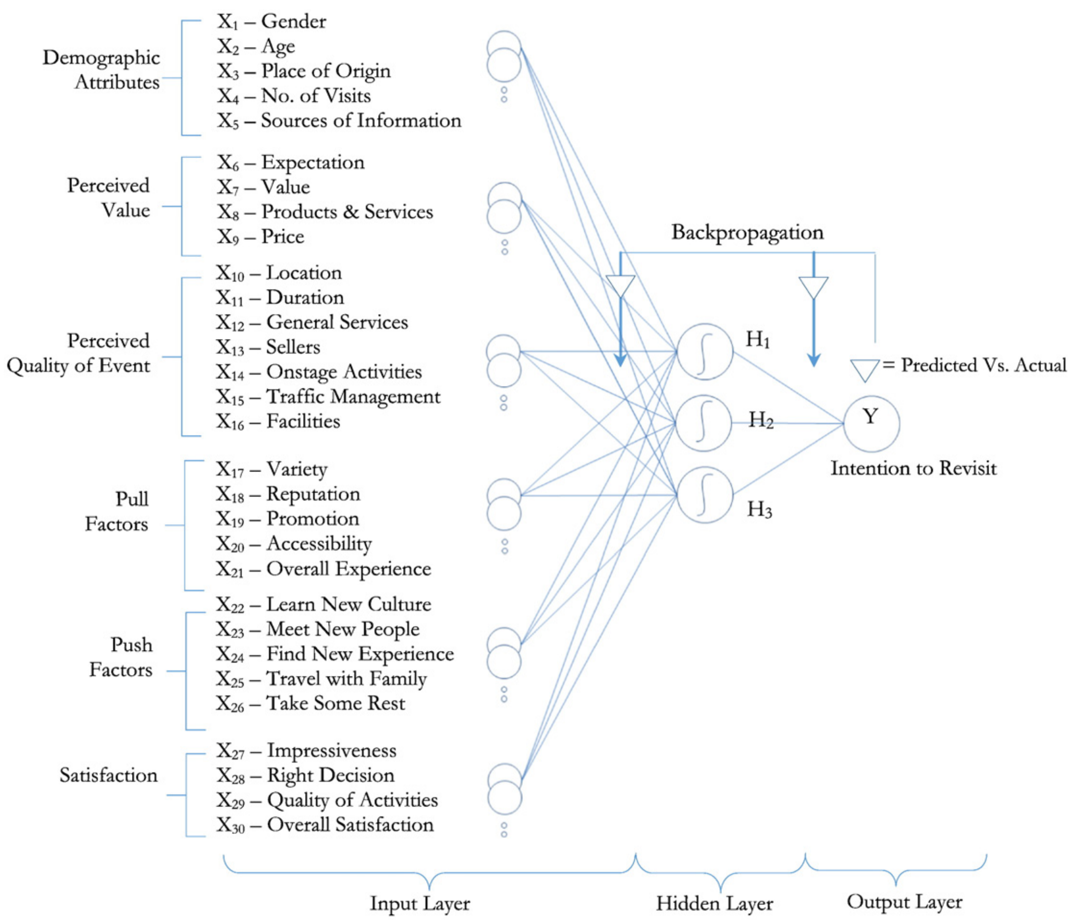


Fig. 5. Cluster analysis.



$$\text{Logit}\ (Y = 1) = -10.138 - 20.902H_1 - 20.872H_2 - 11.706H_3\ [\text{where Intention to Revisit (Y): Yes (Y = 1) and No (Y=0)}]$$
$$H_1 = \tanh(-10.713 - 20.412X_{20} - 3.596X_6 + 5.787X_{16} + 2.99X_{12} + 1.91X_{11} - 2.68X_{24} + \ldots + 1.16X_9)$$
$$H_2 = \tanh(1.226 - 3.181X_{20} + 2.638X_6 - 2.119X_{16} - 3.035X_{12} - 5.762X_{11} - 3.708X_{24} + \ldots + 1.927X_9)$$
$$H_3 = \tanh(-5.897 + 3.55X_{20} - 9.808X_6 - 7.462X_{16} + 5.742X_{12} + 4.733X_{11} - 2.823X_{24} + \ldots + 3.279X_9)$$

Fig. 6. Neural network model.

than 3.5 *AND* "Variety" is less than 3.5 *THEN* the probability of returning to the festival is 6.1%
- *IF* "Overall Satisfaction" is less than 3.5 *AND* "Reputation" is greater than 3.5 *AND* "Variety" is greater than 3.5 *THEN* the probability of returning to the festival is 55.6%

The variables selected in this decision tree model are quite in line with those in the neural network model. Overall Satisfaction (X_{30}), Accessibility (X_{20}), Place of Origin (X_3), Variety (X_{17}), and Reputation (X_{18}) are of concern in the rule-based algorithms. Thus, closer attention to these factors can be promptly initiated. For instance, to increase the probability of visitors returning to the festival, the organizer should focus primarily on increasing the variety of products and services along with maintaining the festival's reputation and attendees' overall satisfaction.

5.8.2.3. Stepwise logistic and polynomial regression. For stepwise logistic and polynomial regression, of the 30 variables used in the intention-to-revisit prediction model, only 5 are selected. This finding helps the festival organizers prioritize the important factors associated with the decision to revisit the site. The organizers definitely start by maintaining the festival reputation (X_{18}) for both products and services so that the overall satisfaction (X_{30}) of visitors is always at the excellent level. Interestingly, the other three reasons that influence the decision to revisit the festival include Promotion (X_{19}), Accessibility (X_{20}), and Learn New Culture (X_{22}). Promoting the cultural themes of the festival along with the various promotions for food and products at the site while increasing accessibility to the festival site can increase the likelihood to revisit the site.

However, conventional regression analysis may not be able to explain the visitors' perceptions and behaviors, which is a sophisticated and dynamic nonlinear phenomenon. Thus, the more complex polynomial logistic regression is applied to improve the accuracy of the predicted outcomes and to test whether the interaction effects among variables impact the intention to revisit the site. The results are quite reasonable, as the misclassification rate decreases from 17.03% to 12.62%. In addition to the variables specified in the stepwise logistic regression in Equation #1, the polynomial regression equation (as presented in Fig. 7) indicates that visitors who attend the site to have new experiences (X_{24}), have high expectations of the festival (X_6), and whose overall experience at the festival (X_{21}) is rated very high are likely to attend the festival again. The Location (X_{10}) of the festival site is still an important factor in the tourists' selection of the festival.

On the other hand, Promotion (X_{19}), advertising as a part of marketing campaigns to persuade a target group of tourists, is commonly considered a pull factor for attendees at the festival. Both logistic and polynomial regressions, Fig. 7, imply that the more products and services are promised, the lower the chance attendees intend to return in the future. The finding

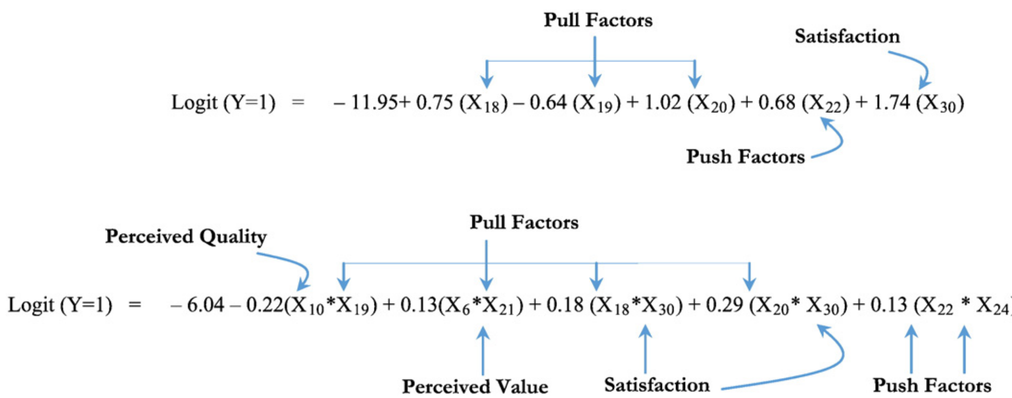


Fig. 7. Stepwise logistic and polynomial regression.

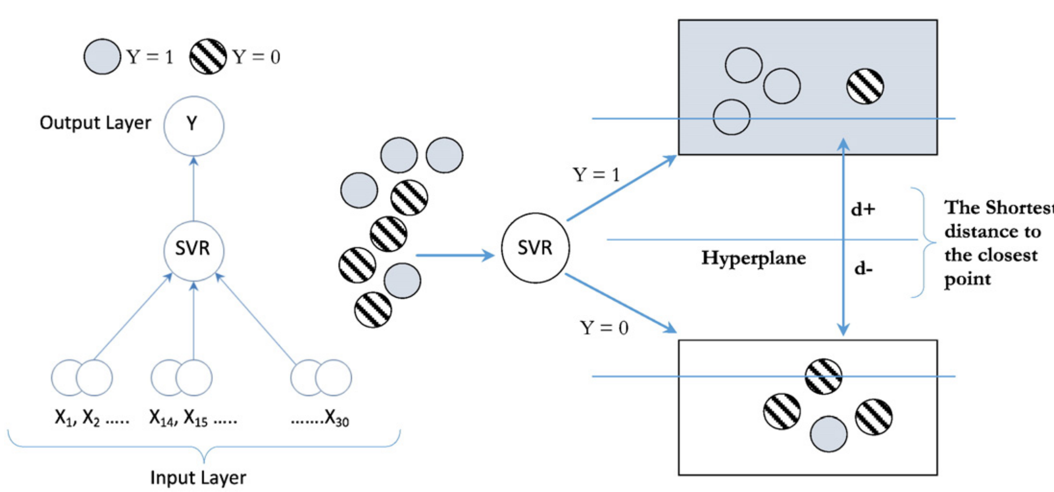
is reasonable, especially when the promotion is exaggerated and the quality of products or services attendees experience is below their expectations.

5.8.2.4. Support Vector Machines (SVMs). Fig. 8 presents the basic architecture of a SVM, which produces a binary classifier of attendees: those who do and those who do not plan to revisit the festival site in the future. The overall accuracy of the model prediction is 87.70% with a false negative of 11.81%. Following are examples of the characteristics of two attendees with the highest probability (greater than 80%) of returning to the site.

- Attendee #1: A local female resident between 40 and 59 years old. She has visited the festival more than once. The main purpose of her visit is to experience a new culture with her family. She receives the information about the festival from cable TV. Although the price is a little high based on her perceived value of her experiences and the variety of products is below her expectations, she is still satisfied with the quality of products and services offered at the site. The overall experience and the impressiveness of the festival overcome the minor issues of traffic management and accessibility to the site.
- Attendee #2: A female visitor who lives in a nearby province and is between 20 and 39 years old. She has also visited the festival more than once. She really likes to take a break from her routine job to spend time with her family. She feels that the facilities such as the rest area and restrooms are not appropriate and need significant improvement. However, the on-stage activities and the variety of products and services far exceed her expectations. She also notes that the location and the timing of the festival are mediocre and the festival is not widely promoted.

5.9. Step# 9: monitoring, measuring, and comparing key performance indicators

The key findings help the festival organizers prioritize the important factors associated with the intention to revisit the festival site in the



The SVM Procedure

Support Vector Classification	C_CLAS
Kernel Function	Linear
Estimation Method	DQP
Maximum QP Size	100
Number of Observations (Train)	317
Number of Effects	41
Regularization Parameter C	0.550000
Classification Error (Training)	39.000000
Objective Function	-55.396654
L1 Loss	1.207368E-14
Inf. Norm of Gradient	1.931788E-14

Squared Euclidean Norm of w	6.462817
Geometric Margin	0.786718
Number of Support Vectors	125
Number of S. Vector on Margin	93
Norm of Longest Vector	6.403124
Radius of Sphere around SV	6.020797
Estimated VC Dim of Classifier	235.277105
Linear Kernel Constant (Fit)	-1.173039
Linear Kernel Constant (PCE)	-1.173039
Number of Kernel Calls	773644

Fig. 8. Basic architecture of SVM model.

future. This study shows that advanced business analytics through data mining approaches (cluster analysis and predictive modeling) are capable of understanding visitors' profiles, given sufficient data with the proper input variables. Thus, closer attention to those factors can be promptly initiated to increase the probability that visitors will return to the festival.

6. Managerial and practical contributions

While the concepts of business intelligence and business analytics have been adopted as a new research paradigm in many disciplines, we have seen very few applications in the tourism industry that fully explore its capabilities. Specifically, much attention has been paid to the value that tourism planners, festival organizers, and marketing analysts at the festival sites could create through the use of massive data, business intelligence, and analytics, which have helped top-performing organizations successfully offer visitors an improved festival experiences and promote follow-up campaigns to retain visitors for future events.

Additionally, while the need for advanced predictive analytics to analyze ample data for understanding visitors' satisfaction with their travel experiences and for timely decision making is apparent, such analytics are not available in traditional statistical analysis methods such as conventional, spreadsheet-based, regression modeling, crosstab analysis, and descriptive statistics.

The uniqueness of this study is the business intelligence framework that is proposed in order to gain insight into the behavior of tourists visiting destination sites. In particular, successful festival organizers require carefully planned and comprehensive analyses of data and information obtained from tourists, both from those who plan to revisit the events and from those who do not. The ability to identify and understand attendees' characteristics through the data mining approach has become a necessity for attracting and retaining the most valuable tourists for the event. Although this study originated within the context of a specific local festival in Thailand, the challenges of seeking repeat travelers to destinations or events in increasingly competitive market environments are quite typical of what is observed for any other travel destination. The framework presented in this study can provide fruitful avenues for exploring the implications of business intelligence and business analytics in any tourism industry, contributing to both knowledge and practice.

7. Challenges and lessons learned

How can we effectively design a BI solution for the problem of understanding why people do or do not intend to return to the festival? The first challenge is to set up the right goals and objectives and to identify the project timeline within the budget in order to meet the expectations and requirements of the festival organizers. Unrealistically high expectations always lower the perceived benefits business intelligence can deliver. The second challenge for business intelligence initiatives is the data integration for model development in Step 4. Thus, approximately 60% of the project time is devoted to preparing, integrating, and cleaning the data to ensure its quality before applying data analytics. The third challenge is in Step 7 – Creating a Data Analysis Framework, – where transforming the focus from descriptive analysis to advanced predictive modeling requires time and resources from experts or staff with strong technological skills. Marketing analysts and festival organizers are more comfortable with the data visualization and exploration in Steps 5 and 6 and often try to skip Steps 7 and 8 in the BI framework. Thus, it is important to empower both marketing analysts and festival organizers to experiment with analytical tools in order to develop analytical and technological skill sets so that they can exploit the full spectrum of business intelligence and analytics techniques. Last, even though the proposed framework and the results of this study help create awareness of BI initiatives, the attitude toward adopting such advanced technology and concepts is still not positive and not all associates are fully engaged in experiments with this BI project. Only certain assigned groups are proactive and try to collect feedback from peers for improvement, which is why the sponsors of the BI project inevitably should come from top management (the local government in this study), who organizes the festival event and recognizes the value of what BI can do.

8. Conclusion

This article provides a short introduction to the use of business intelligence (BI) in the tourism industry. It contains a brief description of a BI framework that has been widely applied in other fields but is still relatively novel in tourism. The proposed framework includes 1) setting up business goals and objectives, 2) defining key performance indicators, 3) collecting data pertinent to the business domain, 4) integrating the data for model development, 5) making sense of festival destination and visitors attending the site, 6) gaining insights into visitors' behavior through data exploration, 7) creating a data analysis framework/scenario, 8) turning knowledge from the data analysis into actionable plans, and 9) monitoring, measuring, and comparing key performance indicators. The case presented in this study has elaborated on a real-world application in festival tourism. Festival organizers are looking for ways not only to satisfy festival visitors so that they become repeat visitors, but also to gain new visitors based on advertising, promotion, or positive word-of-mouth to family and friends from current visitors. Any knowledge and key findings from the data analysis through data exploration and predictive modeling in this business intelligence framework can be used to provide valuable feedback to the festival organizers so that any follow-up marketing campaign to retain the visitors for future events or actionable plans to improve visitors' satisfaction can be promoted.

References

Bloom, J. Z. (2005). Market segmentation: A neural network application. *Annals of Tourism Research*, *32*(1), 93–111.

Byrd, E. T., & Gustke, L. (2007). Using decision trees to identify tourism stakeholders: The case of two Eastern North Carolina counties. *Tourism and Hospitality Research*, *7*(3–4), 176–193.

Byrd, E. T., & Gustke, L. (2011). Using decision trees to identify tourism stakeholders. *Journal of place management and development*, *4*(2), 148–168.

Chase, C. W., Jr. (2009). *Demand-driven forecasting: A structured approach to forecasting.* Hoboken, NJ: John Wiley & Sons.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Ciflikli, C., & Kahya-Özyirmidokuz, E. (2010). Implementing a data mining solution for enhancing carpet manufacturing productivity. *Knowledge-Based Systems*, *23*(8), 783–788.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, *84*(1), 98–107.

Davenport, T. H. (2010). The new world of business analytics. *International Institute for Analytics on-line* (Retrieved March 14, 2016, from) http://www.sas.com/resources/asset/IIA_NewWorldofBusinessAnalytics_March2010.pdf.

Dhond, A., Gupta, A., & Vadhavkar, S. (2000). *Data mining techniques for optimizing inventories for electronic commerce*Paper presented at the Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.

Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework for tourism sustainability: A business intelligence application from Sweden. *Turizam: znanstveno-stručni časopis*, *61*(2), 121–148.

Getz, D. (2008). Event tourism: Definition, evolution, and research. *Tourism Management*, *29*(3), 403–428.

Golmohammadi, A., Shams Ghareneh, N., Keramati, A., & Jahandideh, B. (2011). Importance analysis of travel attributes using a rough set-based neural network: The case of Iranian tourism industry. *Journal of Hospitality and Tourism Technology*, *2*(2), 155–171.

Harding, J., Kusiak, A., & Shahbaz, M. (2006). Data mining in manufacturing: A review. *Journal of Manufacturing Science and Engineering*, *128*(4), 969–976.

Jackson, J. (2002). Data mining: A conceptual overview. *Communications of the Association for Information Systems*, *8*, 267–296.

Jang, S. S., & Feng, R. (2007). Temporal destination revisit intention: The effects of novelty seeking and satisfaction. *Tourism Management*, *28*(2), 580–590.

Kim, J., Wei, S., & Ruys, H. (2003). Segmenting the market of West Australian senior tourists using an artificial neural network. *Tourism Management*, *24*(1), 25–34.

Kuo, R. J., Akbaria, K., & Subroto, B. (2012). Application of particle swarm optimization and perceptual map to tourist market segmentation. *Expert Systems with Applications*, *39*(10), 8726–8735.

Kusiak, A. (2006). Data mining: Manufacturing and service applications. *International Journal of Production Research*, *44*(18–19), 4175–4191.

Law, R., Leung, R., & Buhalis, D. (2009). Information technology applications in hospitality and tourism: A review of publications from 2005 to 2007. *Journal of Travel & Tourism Marketing*, *26*(5–6), 599–623.
Li, X. R., Cheng, C. K., Kim, H., & Petrick, J. F. (2008). A systematic comparison of first-time and repeat visitors via a two-phase online survey. *Tourism Management*, *29*(2), 278–293.
Lin, C. T., & Huang, Y. L. (2009). Mining tourist imagery to construct destination image position model. *Expert Systems with Applications*, *36*(2), 2513–2524.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* New York, NY: Houghton Mifflin Harcourt.
Mikulić, J., Paunović, Z., & Prebežac, D. (2012). An extended neural network-based importance-performance analysis for enhancing wine fair experience. *Journal of Travel & Tourism Marketing*, *29*(8), 744–759.
Minelli, M., Chambers, M., & Dhiraj, A. (2012). *Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses.* Hoboken, NJ: John Wiley & Sons.
Narasimhan, R., Swink, M., & Kim, S. W. (2005). An exploratory study of manufacturing practice and performance interrelationships: Implications for capability progression. *International Journal of Operations & Production Management*, *25*(10), 1013–1033.
Park, D. B., & Yoon, Y. S. (2009). Segmentation by motivation in rural tourism: A Korean case study. *Tourism Management*, *30*(1), 99–108.
Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. *Communications of the Association for Information Systems*, *37*(1), 448–472.
Pyo, S., Uysal, M., & Chang, H. (2002). Knowledge discovery in database for tourist destinations. *Journal of Travel Research*, *40*(4), 374–384.
Quintal, V. A., & Polczynski, A. (2010). Factors influencing tourists' revisit intentions. *Asia Pacific Journal of Marketing and Logistics*, *22*(4), 554–578.
Rivera, J., & Meulen, R. (2014). Gartner says advanced analytics is a top business priority. (Retrieved Feb 22, 2014, from) http://www.gartner.com/newsroom/id/2881218.
SAS (2012). Big data meets big data analytics. (Retrieved Feb 22, 2015, from) http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/big-data-meets-big-data-analytics-105777.pdf.
Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligence systems.* Upper Saddle River, New Jersey: Pearson Education.
Um, S., Chon, K., & Ro, Y. (2006). Antecedents of revisit intention. *Annals of Tourism Research*, *33*(4), 1141–1158.
Wang, K. (2007). Applying data mining to manufacturing: The nature and implications. *Journal of Intelligent Manufacturing*, *18*(4), 487–495.

Thanathorn Vajirakachorn is a lecturer in the Department of Tourism Management, School of Business at the University of the Thai Chamber of Commerce, Bangkok, Thailand. She received her M.S. degree in Hospitality and Tourism from University of Wisconsin-Stout, and Ph.D. in Tourism from Texas A&M University. Her major research interests include community-based tourism, sustainable tourism, and gastronomic tourism.

Jongsawas Chongwatpol is an assistant professor in NIDA Business School at the National Institute of Development Administration. He received his B.E. in industrial engineering from Thammasat University, Bangkok, Thailand, two M.S. degrees in risk control management and management technology from the University of Wisconsin - Stout, and a Ph.D. in business administration with a specialization in management science and information systems from Oklahoma State University. His research has recently been published in major journals such as *Decision Support Systems, Decision Sciences, European Journal of Operational Research, Energy, Industrial Management and Data Systems, Journal of Information Technology Teaching Cases, and Journal of Business Ethics*. His major research interests include business intelligence, decision support systems, big data, data mining, RFID, manufacturing management, and supply chain management.