

Data Mining in Tourism Data Analysis: Inbound Visitors to Japan

4-3-2017

Valeriya Shapoval

University of Central Florida, valeriya.shapoval@ucf.edu

Morgan C. Wang

University of Central Florida, Chung-Ching.Wang@ucf.edu

Tadayuki Hara

University of Central Florida, tadayuki.hara@ucf.edu

Hideo Shioya

Find similar works at: <https://stars.library.ucf.edu/rosenscholar>

University of Central Florida Libraries <http://library.ucf.edu>

 Part of the [Hospitality Administration and Management Commons](#), and the [Tourism and Travel Commons](#)

Original Citation

Shapoval V., Wang, C. Hara T., & Shioya, H. (2017) Data mining in tourism data analysis: Inbound visitors to Japan. *Journal of Travel Research*.

This Paper is brought to you for free and open access by the Rosen College of Hospitality Management at STARS. It has been accepted for inclusion in Rosen Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.

Data Mining in Tourism Data Analysis: Inbound Visitors to Japan

Abstract

The increasing power of technology puts new, advanced statistical tools at the disposal of researchers. This is one of the first research papers to use a data mining tool, namely, decision trees, to analyze the behavior of inbound tourists for the purposes of effective future destination marketing in Japan. The research results of approximately four thousand observations show the main motivation for visitors' future return is not driven by experiences had during their most current visit but rather by anticipated experiences in the future, such as experiencing hot springs or immersing themselves in beautiful natural settings. The data mining method largely excludes the possibility of the intrusion of researcher subjectivity and is conducive to useful discoveries of certain patterns with large visitor data sets, providing governments and destination marketing organizations with additional tools to better formulate effective destination marketing strategies.

Keywords: big data analysis, decision trees, quantitative destination marketing, data mining, Japan

Data Mining in Tourism Data Analysis: Inbound Visitors to Japan

Growing international tourism and travel has increased competition among tourism destinations throughout the world. Destinations are a combination of tourist products and services that create an integral experience for tourists and are consumed under the brand name of the destination (Leiper, 1995; Buhalis, 2000). Historically, destinations have been considered to be specific geographical locations, such as cities or countries (Hall, 2000). On the other hand, there is a new trend in defining a destination as a concept that can be subjectively interpreted by

tourists based on their purpose, culture, past experiences, etc. (Buhalis, 2000). It is increasingly recognized that the core of any destination's successful performance is determined by satisfied tourists who intend to return in the future and who will recommend this destination to their friends and families (Chi & Qu, 2008; Assaker, Vinzi, & O'Connor, 2011; Valle et al., 2006). Ryan (1991) argued that if the tourism industry is to continue satisfying tourists, it has to adopt societal marketing strategies, carefully monitor tourist satisfaction, and use the information collected to create success. Thus, understanding which attributes of a given destination create tourist satisfaction as well as the types of tourists who are willing to come back in the future enables destination marketing organizations (DMOs) to plan future destination marketing strategies that could be essential for an emerging tourist destination such as Japan.

For the past several decades, Japan has established its image as an industrial country, but with increasing competition from other countries using its manufacturing-led growth model, it has also displayed an increasing interest in inbound tourism and has pursued several marketing campaigns to increase global awareness of Japan as a tourism destination. According to the Japan National Tourism Organization (JNTO), *Japan witnessed a historical record of 19,737,409 inbound visitors in 2015, and in 2016, the Japanese government announced an ambitious plan to increase the number of annual inbound visitors to Japan to 40 million by 2020, and even to 60 million by 2030. (Japan National Tourism Organization, 2016).*

As tourism becomes increasingly important to Japan as a destination, academic research is gradually starting to address questions of motivation, intention to return, positive word of mouth, and satisfaction related to Japan as a tourism destination. However, only a limited amount of research has been carried out regarding international tourism in Japan, and that is mainly qualitative in nature and includes very little empirical research. By contrast, extensive

contributions that employ various qualitative and quantitative methodologies applied to research on customer satisfaction and behavioral intentions for specific destinations have enriched the current body of literature (Buhalis, 1999; Pearce 2014; Som & Badarneh, 2011; Okamura & Fukushige, 2010; Uzuma, 2009; Park & Gretzel, 2007; Liu, Sigaw, & Enz, 2008; Pizam, Neumann, & Reichel, 1978). The existing literature is nevertheless not comprehensive, and technological advances have led to methodological and statistical advancements that provide opportunities to take a broader view on tourism behavior for destination marketing by using data mining techniques (Goh, Law, & Mok, 2008).

Data mining techniques are extremely underutilized in tourism research, and empirical research on Japan as a tourism destination using data mining techniques is very limited. Hence, the purpose of this research is twofold: first, to provide an overview of data mining and its potential in tourism research using decision trees; and second, to fill the empirical research gap in the tourism literature related to Japan as a tourist destination.

This article, which presents findings related to tourist satisfaction and intention to return to Japan drawn from research using advanced statistical methods of data mining, has three specific objectives: 1) to identify the most important experiences of inbound tourists to Japan as a destination; 2) to identify the preferences, likes, and dislikes of the tourists; and 3) to identify how those experiences and preferences affect satisfaction and future intention to return to Japan as a destination choice.

Theoretical Background

Data Mining

Advances in technology and computer power have made it possible to collect immense amounts of data across many different fields. There is an increasing need for tools that will assist

in extracting useful information from the growing amounts of data and turn it into knowledge (Fayyad, Piateetsky-Shapiro, & Smyth, 1996). Many industries are enhancing their competitiveness by adopting data mining technology for various purposes: gauging customer preferences in e-commerce and retail; ascertaining medical history in health care; assessing risk factors in insurance; and gathering financial data in banking, to name just a few. The nature of the tourism and hospitality sector has made it one of the largest users of informational technology (Sheldon, 1994; Buhalis, 1998). Information about tourists is being accumulated at an increasing pace, and it is getting progressively more difficult for destinations to stay competitive, to increase their market share, and to maintain a competitive advantage. Destination management organizations will find a growing need to use data mining if they wish to stay competitive (Pyo, Uysal, & Chang, 2002).

With an acceptably accurate learning model, one cannot only understand but also predict expected values in the tourism industry. For example, a tourism agency may choose to use its visitor database to predict future arrivals and patterns of consumption. Given an updated visitor profile, agencies will be more prepared, based on actual visitor behavior, to: 1) meet the needs of visitors with better marketing material; 2) establish necessary collaborations with agencies such as transportation and lodging; 3) share information and work with other destinations within a country to improve the country's desirability for future touristic visits (Apte & Weiss, 1997). While some academic articles (Buhalis, 1997; Pyo, Uysal, & Chang, 2002; Buhalis & Law, 2008) have emphasized the need for data mining in tourism rather than empirical research with classical statistics, empirical research undertaken with data mining tools remains highly limited (Kim, Timothy, & Hwang, 2011). **A potential question could be raised: what is the advantage of data-mining-based research compared to already established techniques using classical statistics?**

While both techniques have their advantages and disadvantages, the authors thought it would be prudent to give a brief comparison of the two techniques and provide a rationale for why data mining would be advantageous in this case.

Data mining versus classical statistics.

Managing very large data requires skills that are different from those used in classical statistical analysis. Data mining manages such problems by efficient summaries of large amounts of data, identifying patterns and relationships of past data, and constructing predictors for the future. Classical statisticians have well-established tools for such things. Many statistical models have been utilized for explaining relationships and patterns within given data, and it is therefore tempting to think of data mining as an extended branch of statistics. However, data mining has its own merits, being capable of working with large-scale data sets than the data sets used in classical statistics. Comparatively, there are differences in the approaches to modeling, in which data mining pays less attention to the large-sample asymptotic properties of its inferences and more to the “learning,” including the complexity of the modeling and computation required by large data sets. Of course, both classical statistics and data mining are similar in that they draw inferences from data. However, unlike classical statistics, data mining is more tolerant toward discrete-valued variables and seeks to minimize a loss function expressed in terms of predictor error, where minimization is achieved by cross-validation (Hosking, Pednault, & Sudan 1997).

One of the oldest definitions of data mining is “the non-trivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley et al., 1992). Data mining uses machine learning algorithms to find patterns of relationships between data elements in large, noisy, and messy data sets, thereby facilitating actions that enhance in some form (diagnosis, profit, detection, etc.) knowledge discovery in that data (Nisbet, Edler, & Miner,

2009, p. 17). As data mining evolved, a new definition was proposed as follows: “Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potential, useful, and ultimately understandable patterns in data (Fayyad et al., 1996).

One big difference between classical statistics and data mining is that classical statistics has large subjective components, known as predictive models, the main goal of which is to estimate parameters and/or confirm or reject hypotheses. On the other hand, from a data mining mindset, the correct model is unknown. In fact, the goal of the analysis is to discover the correct model even if it is not correct. In classical statistics, models must be specified, whereas in data mining, a series of competing models will be specified and selected based on data examination. This preferential ordering addresses the issue of overfitting. There are many other things which can be said about the differences between classical statistics and data mining techniques; however, this is not the purpose of this paper. In summary, one can say that statistical learning (data mining) is much more manageable when there are no restrictions placed on the model for a given data set, in other words, where analyses are data driven and the complexities of the given machine learning algorithms are dependent on the underlying distribution we desire to learn (Hosking, Pednault, & Sudan, 1997).

An extensive number of data mining techniques have evolved over the years, including but not limited to: decision trees, neural networks, regression analysis, text mining, association rules, and clustering.

Data preparation and reduction are essential steps in data mining. Unlike data sets used in classical statistics, it is impossible to “eyeball” data mining datasets where variables could be counted in hundreds and observations in millions because, just as in classical statistics, the quality of the prediction and accuracy of a model depend on the quality of the data. Furthermore,

variables should be reduced and manipulated into analytical data sets. Finally, once a data set is cleaned and finalized, similar to classic statistics, an appropriate statistical tool is chosen for data analysis, such as neural networks, time series, decision trees, etc.

Destination Marketing

Tourism is one of the world's major industries that contributes significantly to the global economy and has become one of the major sources of wealth for a number of developing and developed countries. Tourism takes place at destinations; consequently, a destination is taken as the fundamental unit of analysis (WTO, 2002). Destinations are also a focal point of destination marketing, an essential tool of tourism destinations in an increasingly globalized and international tourism market (UNWTO, 2011). Destinations are a conglomeration of tourist service and experiences (Buhalis, 2000). Understanding tourists' perceptions is essential to a successful tourism destination because they influence a tourist's choice of destination (Ahmed, 1991), their satisfaction, and their decision whether to return (Weiermair, 2000). The increasing competition among tourist destinations in the last several decades has prompted concern among destination marketing managers and industry practitioners about perceptions of a destination by tourists (Wang & Pizam, 2011). The marketability of destinations as well as offered services, entertainment, lodging, transportation, and shopping leave an impression on visitors, their sense of satisfaction, and their decision whether to come back in the future. Thus, the following questions are raised: How can a DMO best communicate with stakeholders and the market? How can a DMO engage with visitors to stimulate repeat visits? Finally, how can a DMO filter the vast amount of information to obtain a set of manageable rules to predict visitor behavior and ensure their satisfaction and loyalty? (Pike, 2012).

Destination marketers do extensive research to identify prospective visitors who have not visited yet (suppressed demand) and potential tourists (active demand) (Athiyaman, 1997). Destination marketing organizations need to know how their destination is perceived by potential visitors in order to better target their market, develop more appropriate tourism products, and increase destination attractiveness (Phillips & Back, 2011). For example, cultural differences, the extent of planning time before vacation, and the number of people in the group influence tourist expenditure (Leasser & Dolnicar, 2012). A review of past literature shows an increasing number of articles that deal with aspects of destination marketing, customer satisfaction, and behavioral intentions in tourism overall and for a specific destination. For instance, Kozak and Rimmington (2000) looked at tourist satisfaction in Mallorca, Spain. Baloglu and McCleary (1999) looked at U.S. international pleasure travelers to four Mediterranean destinations. Yoon and Uysal (2005) studied motivation and satisfaction of tourists in Northern Cyprus. Compelo, Aitken, and Gnoth (2011) looked at visual rhetoric in the destination marketing of New Zealand. Finally, Dwyer et al. (2014) studied destination marketing and return on investment in Australia.

Inquiring into tourist perception of a destination is generally aimed at looking into customer satisfaction and intention to return. Literature relating to measuring the two can be successfully arranged into two groups (Hallowell, 1996). Service management literature postulates that customer satisfaction leads to customer loyalty and subsequently leads to profitability (Hallowell, 1996; Reinartz & Kumar, 2002). Marketing literature claims that if customers are happy with a product, they purchase it again and tell their friends and relatives about it (Maxham, 2001; Ranaweera & Prabhu, 2003; Brown et al., 2005). Similarly, this concept could be applied to the body of tourism literature, which finds a significant correlation between satisfaction and future intention to return (Gallarza & Saura, 2006; Hernandez-Lobato,

Solis-Radilla, & Moliner-Tena, 2006). A number of articles have examined differences between first-time and repeat visitors (Woodside & Lysonski, 1987; Lupton, 1997; Okamura & Fukushige, 2010; Fuchs & Reichel, 2011) and have established that repeat visitors are more likely to choose the same destination. First-timers will reduce their stereotypes and obtain a better and deeper understanding of a destination (Pool, 1965). Repeaters will move beyond simple stereotyping and build a more subtle and complex understanding. (Fakeye & Crompton, 1991; Mishler, 1965). This of course happens when sufficient time has been spent at a destination and the tourist has had sufficient saturation through establishing different contacts and relationships (Mishler, 1965).

It is generally accepted that satisfaction is essential for destinations to have repeat visitors and intention to return to a destination will depend on the level of satisfaction visitors had with products and services. On the other hand, if visitors are not satisfied, they will be less likely to return to a destination.

Since this study is looking into destination experiences and attributes, we will use a definition of tourist satisfaction proposed by Pizam et al. (1978): tourist satisfaction is the result of interaction between a tourist's experience at the destination and the expectations he or she had about that destination.

While it is true that satisfaction and intention to return are highly important for tourism, destinations are an amalgam of tourism products and services that create experiences for the consumer. There is a plethora of important dimensions that could potentially contribute to consumer satisfaction and subsequent return. There is also an increasing trend in the recent stream of research that reveals different dimensions that influence tourists' destination perceptions, satisfaction, and loyalty (table 1).

Table 1 is here

For example, Yoon and Uysal (2005) examined “push and pull” motivation factors for satisfaction and destination loyalty. “Push” motivation factors included relaxation, family togetherness, safety, and fun. “Pull” motivators included weather, shopping, cleanliness, night life, and local cuisine. “Push” motivators had a significant impact on destination loyalty, and satisfaction with destination leads to destination loyalty. Lee, Kyle, and Scott (2012) approached destination loyalty from an events perspective. They found that satisfaction with a special event (a festival) led to destination preference and place attachment (place identity and dependence). B., Lee, C., Lee, and J., Lee (2014) studied change in tourist perceptions of destination image before and after a trip, and the destination image was found to have been significantly impacted by satisfaction. Destination image dimensions such as amenity and hygiene, attractions, and accessibility were different for tourists after visiting a destination. Bajs (2015) evaluated the effects of quality of touristic services, destination appearance, emotional experience, and monetary and non-monetary costs on perceived value and subsequently on satisfaction and behavioral intentions. Joppe, Martin, and Wallen (2001) studied Toronto visitors’ perceptions of product and service attributes such as hospitality, accommodations safety, cuisine, and family orientation using an importance satisfaction model. They found that food, accommodations, and sightseeing ranked as very important and excellent on their importance-satisfaction grid, while family-orientation was unimportant and unsatisfactory. Alegre and Garau (2011) found that cuisine, budget, cleanliness, climate, scenery, and access were of explicit importance to tourist satisfaction and destination competitiveness. Chi and Qu (2008) applied an empirical integrative approach to understanding destination loyalty and used destination image, overall satisfaction, and tourist attributes. Ramseook-Munhurrin, Seebaluck, and Naidoo (2014) used tourist

perception of destination image, perceived value, tourist satisfaction, and loyalty for destination marketing for the small island destination Mauritius. In this study, the authors looked at dimensions such as travel environment, attractions, events, infrastructure, sports activities, and perceived value as antecedents of satisfaction and loyalty. While only satisfaction had an impact on loyalty, perceived value and destination image had an impact on tourist satisfaction. Finally, Ozdemir and Simsek (2015) analyzed perceptions of quality, price, and value on satisfaction and destination image. They found that perceived price and quality have a significant impact on destination image.

Another recent trend has been the increase in destination marketing research that takes into consideration specific destination attributes and their effects on tourist satisfaction and behavioral intentions. Research using advanced statistical tools, however, appears to be limited. The authors of the present study took an extensive look at underlying dimensions of “destination image” and “attribute satisfaction.” Destination image includes dimensions such as travel environment, natural attractions, entertainment and events, infrastructure, relaxation, outdoor activities, price, and value while attribute satisfaction includes shopping, activities, lodging accessibility, attraction, dining, and environment. Destination image and attribute satisfaction had a significant impact on overall satisfaction and destination loyalty

Advances in technology have increased researchers’ ability to collect, store, and run calculations on very large data sets (Pyo, Usal, & Chang, 2002). Big data analysis is slowly making progress as a valid research tool in broader social science fields. For example, in the medical field, Qu et al. (2002) used decision trees in their evaluation of a proteomic approach to the simultaneous detection and analysis of multiple proteins for differentiation of prostate cancer patients from non-cancer patients. DeReyck, Degraeve, and Vandenborre (2008) used decision

trees for evaluation as an alternative approach to valuing real options based on a certainty-equivalent version of the net present value formula. Goh, Law, and Mok (2008) incorporated rough sets theory into tourism demand analysis and created a tourism climatic index by using data mining techniques. They found that climate and leisure time have a stronger impact on tourist arrivals than economic factors. Wicker and Breuer (2013) used decision trees to evaluate organizational problems for recruitment/retention of members at non-profit sports clubs. Duncan (1980) used decision trees in his evaluation of organizational structure and design. Min, Min and Emam (2002) used data mining approach in developing hotel customers' profiles. Chang, Chen, Kuo, Hsu and Cheng (2016) applied data mining techniques (decision trees) for tourist loyalty intentions in hotel sector. Specifically, authors were looking at hotels/ physical environment and social interaction to gauge a customer loyalty. Kim, Timothy, and Hwang (2011) used DT analysis to evaluate Japanese tourists' shopping preferences and intention to revisit Korea.

Finally, this paper is making one of the first attempts to use DTs to analyze tourist satisfaction and intention to return by utilizing large data sets on inbound visitors using destination-specific attributes of Japan as a destination.

Japan as a destination

Japan is still largely undiscovered by mass tourism. Mainly known for its industrial power, Japan as a tourism destination is still overshadowed by its industrial and business image. Even though a limited amount of academic research on Japan as a destination does indeed point to its tremendous potential as a tourism destination, this potential remains generally untapped.

Nevertheless, research on inbound tourism to Japan remains highly limited. According to Uzama (2009), the Japanese marketing campaign "Yokoso! Japan" was mainly unsuccessful in advertising Japan as a desirable tourism destination, and in spite of government interest in

promoting Japan, it did not go beyond simple promotion. Omura and Fukushige's (2010) research on international tourists to Japan looked into the differences between first-time visitors and repeat visitors to the Kansai area of Japan and found that first-time visitors were interested in sightseeing while repeat tourists were more involved and interested in participating in events. Such relatively limited existing research emphasizes the need for additional empirical research about Japan as a destination for the purposes of its destination marketing.

Pyo, Usal and Chang (2002) emphasized a need for data mining analysis and its application to the distribution of knowledge about tourists and destinations as well as market information. Similarly, the authors stressed that promotional activities could be more effective after the characteristics of the destination have been understood and defined. Destinations count their visitors in thousands and millions. DMOs and other government institutions have an extensive amount of data that reflects actual tourists' behavior, but data mining is generally limited to private organizations and consulting firms in the hospitality and tourism industry. Buhalis (2000) emphasized that tourism research is extensively dynamic and continuous research is necessary to follow developments. However, despite the possible benefits that data mining research can provide to destination marketing, empirical research using data mining techniques in the tourism industry has been sorely lacking.

Methodology

Data Mining–Decision Trees

Decision trees (DTs) are a form of multiple variable analysis. A decision tree "...is a structure that can be used to divide up a large collection of records into successively smaller sets of records by applying a sequence of simple decision rules" (Berry & Linoff, 2000, p. 6). Another definition of a DT provided by Nisbet, Edler, and Miner (2009) states that a "DT is a

hierarchical group of relationships organized into tree-like structures, starting with one variable (like a trunk or an oak tree) called a root node” (p. 241). The root node is split into multiple branches using a split criterion. Each split is defined in terms of impurity measure reflecting how uniform resulting cases are. Each split node is referred to as a parent node, and the following splits are called child nodes. Splits continue until the final or terminal node with the minimum number of cases is reached. For example, Figure 1 is a small illustration of decision trees used to indicate patterns of travel behavior based on age, gender, and marital status.

Figure 1 is here

DTs are a very appealing method of analysis for the present study due to their relative power, ease of use, robustness, ability to handle ordinal data (Likert scale), and ease of interpretability. It is a collection of one-cause, one-effect relationships presented in the form of a tree. DTs try to find strong relationships between input and target variable; when a set of values is identified that have a strong relationship to a target, thereafter all of those values are grouped into the bin that forms branches of a DT.

Impurity-based Criterion. In many cases, a DT split is done according to the value of a single variable. The most common criterion for a split would be an impurity-based split as used in this study. The impurity-based criterion is briefly represented as follows:

Given random variable x with k discrete values, distribution according to $P = (p_1, p_2 \dots p_k)$, an impurity measure is a function of $\varphi: [0, 1]^k \rightarrow \mathcal{R}$ that satisfies the following conditions:

$\phi(\mathbf{P}) \geq 0$

$\phi(\mathbf{P})$ is minimum if $\exists i$ such that component $p_i = 1$.

$\phi(\mathbf{P})$ is maximum if $\forall i, 1 \leq i \leq k, p_i = 1/k$.

$\phi(\mathbf{P})$ is symmetric with respect to components of \mathbf{P} .

$\phi(\mathbf{P})$ is smooth (differentiable everywhere) in its range.

Given the training set S , the probability vector of the target attribute y is defined as:

$$P_y(S) = \left(\frac{|\sigma_{y=c_1} S|}{|S|}, \dots, \frac{|\sigma_{y=c_{|dom(y)|}} S|}{|S|} \right)$$

The goodness-of-split due to the attribute a_i is defined as a reduction in impurity of the target attribute after partitioning S according to the values $v_{i,j} \in dom(a_i)$:

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|dom(a_i)|} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}} S))$$

(Rokach & Maimon, 2010, p. 153)

Information Gain. Out of three tests (gini, chi-square, and entropy), the entropy information gain criterion was chosen for the purposes of this study. Information gain is an impurity-based criterion that uses the entropy measure (originating from information theory) as the impurity measure (Quinlan, 1987). Entropy information gain is represented as:

$$\text{InformationGain}(a_i, S) = \text{Entropy}(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \text{Entropy}(y, \sigma_{a_i=v_{i,j}} S)$$

where

$$Entropy(y, S) = \sum_{c_j \in dom(y)} \frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|}$$

(Rokach & Maimon, 2010, p. 153).

Sample and Data Collection.

Data was acquired by the Japan Travel Bureau (JTB) Foundation on behalf of the Japan Tourism Agency in 2010. The JTB Foundation is the largest travel agency in Japan and one of the largest travel agencies in the world that specializes in tourism. The JTB Foundation is a non-profit research organization affiliated with the JTB. (The JTB was established in 1912 and became a for-profit company in 1963.) Data collection was conducted at international airports and seaports in Japan as a part of a tourist expenditure survey series undertaken for the Japan Tourism Agency. Inbound tourists to Japan were approached at random by representatives of the JTB Foundation with an iPad in their hands. Participation in the survey was voluntary, and no monetary incentives were provided for participation. Questions were dictated by the interviewer to the interviewees and answers recorded on the spot, after which the iPad sent the data immediately to the database.

While data mining could potentially offer substantial benefits to research and development, utilizing a large data set potentially raises legal questions and potential liabilities. In 2006, AOL (650 thousand users) and Netflix (100 million ratings) released “anonymized” user data. Potential anonymization failed for both organizations, however, creating legal concerns (Walton, 2014). The authors emphasize that AOL used about 30% of their users’ information (total users is approximately 2.1 million (Pagliery, 2015)) and Netflix released an extensively high number of reviews. The ratio of sample to population of users for that organization was

considerably high, thus creating privacy concerns. Data collected by the JTB Foundation was anonymous and the sample represented less than one percent (1%) of total foreign inbound tourists to Japan (8.65 million total in 2010 (JTB Tourism Research and Consulting, 2016)). Hence, any potential privacy threats have been eliminated.

Data were collected using a Likert and binary scale, and out of a total sample size of 6,000, roughly 4,000 usable observations were obtained. Due to the large sample size, use of classical statistical tools was not appropriate; therefore, the decision tree data mining technique was used for data analysis. Specifically, due to the binary and ordinal scales used in the survey, decision trees with two-step modeling (with two dependent variables) were used to summarize and interpret behavioral and purchasing patterns of inbound tourists in Japan.

In this paper we will use data mining as an exploratory tool and extract hidden knowledge through a set of rules that connects a collection of inputs. In a sense, DTs represent a series of questions, where an answer to a question determines the follow-up question, thereby creating a pattern. The decision tree is probably one of the most popular and powerful techniques used in data mining (Berry & Linoff, 2000). DTs do not have strict assumptions concerning the functional form of the model, but they do have computational efficiency, are robust against outliers, are resistant to the curse of dimensionality, and require less data preparation than other data mining tools.

Measurements.

This study employed a casual research design. The survey questionnaire consisted of the following major sections: tourist attributes of satisfaction, overall satisfaction, intention to return, and demographic questions for tourists requesting information such as country, party size, gender age, and number of children.

Attributes of Satisfaction. Destination response encompassed information about current trips to Japan, purpose of the visit, expenditures, transportation, stay arrangements, shopping, sources of information, activities at destination, satisfaction with Japan as a destination, and intention to return in the future to Japan. The survey consisted of over 150 questions measured on a five-point Likert scale and 0/1 binary responses.

Overall Satisfaction. A single overall measure of satisfaction was used in this study for its ease of use and empirical support. Satisfaction was measured on a seven-point Likert scale with 1 being highly dissatisfied and 7 highly satisfied.

Behavioral Intentions. A single measure for intention to return was used in this study for its ease of use and empirical support. Intention to return was measured on a seven-point Likert scale where 1 indicated definitely not returning and 7 definitely returning.

Results

The top most important variables are listed in Table 1 and Table 2. Out of 150 variables, 18 are represented in the tables. The variables' importance was selected by decision trees and is measured on a continuous scale (decimals) from 1 to 0, with importance decreasing as it approaches 0. The top fifteen variables for satisfaction and intention to return are listed in the tables. Variables are listed in order of importance for a variable satisfaction and future intention to return to Japan as dependent variables:

Table 2 and Table 3 are here

Decision Tree Rules

Important variables provide a snapshot of what is important to tourists when they travel to Japan. Conversely, decision trees provide a deeper understanding by grouping and creating patterns of tourist preferences that provide higher levels of satisfaction and intention to return.

Due to the large number of independent variables, it was not possible to insert a complete decision tree into this paper. However, an excerpt from a decision tree is shown as an example here.

Satisfaction

Figure 2 is here

Intention to return

Figure 3 is here

Demographics of Data

The majority of tourists came from Asian countries (62%) such as Korea (19.51%), Taiwan (18.10 %), and mainland China (14.16%). The second largest group of visitors was from the United States (10.65%). From mainland China, the two largest groups were from Beijing and Shanghai. Gender was rather evenly distributed between men (56%) and women (43%). The average age was 23 years, with a standard deviation of 13 years. The airports that the majority of the tourists arrived at were Narita-Tokyo (53.88%), Kansai-Osaka (17.63%), and New Chitose-Sapporo in Hokkaido (6.212%). The data revealed that 42% of respondents were visiting Japan for the first time, 15% were visiting for the second time, and 10% for the third time. The general distribution of the groups of travelers was: alone (17%), with family (21%), with one or more work colleagues (19%), and with one or more friends (19%). Additionally, 57.9% of the respondents traveled for tourism and leisure, while 25% traveled to participate in business training, conferences, or trade fairs.

Decision Trees

Odds ratios. Odds ratios are used to compare the relative odds of the occurrence of the outcome of interest (e.g., a disease or a disorder) given exposure to the variable of interest (e.g., a health characteristic, an aspect of medical history). An odds ratio is represented by the formula:

$$\text{Odds Ratio } (\Theta) = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}$$

where

OR=1 Exposure does not affect odds of outcome

OR>1 Exposure associated with higher odds of outcome

OR<1 Exposure associated with lower odds of outcome (Bland & Altman, 2000).

Satisfaction. For the purposes of better classification with decision trees, variable satisfaction was recoded into a binary variable where 1 includes highly satisfied and satisfied and 0 includes everything else. That produced binary values that were rather equally distributed between 1 (50.1%) and 0 (48.9%). For the purposes of this study, the top four most important decision tree combinations (rules) were selected. The overall model's misclassification rate is 0.14. The misclassification rate calculates the proportion of an observation being allocated to the incorrect group. It is calculated as follows: Number of Incorrect Classifications/Total Number of Classifications. This indicates an accuracy for the model of 86%.

- Results-Satisfaction.* The odds ratio of tourists being satisfied is higher by
- 2.32 if the tourists are mainly from non-Asian countries, had an experience with Japanese food, paid no higher than \$1,500 for airfare, purchased Japanese fruits, and shopped at a supermarket;

- 2.21 if the tourists are mainly from non-Asian countries, paid no higher than \$1,500 for airfare, experienced Japanese food, stayed less than eight days, and stayed at a Western-style hotel;
- 1.64 if they are from a neighboring Asian country (Korea, China, Taiwan, Hong Kong, or Thailand), stayed at a Japanese-style inn, experienced Japanese food, came for tourism/leisure, incentive travel, study, or international conference, and came through one of the two main airports (Narita/Haneda); and
- 1.51 if the tourists are from a neighboring Asian country (Korea, China, Taiwan, Hong Kong, or Thailand), experienced Japanese food, came for tourism or exhibition/conference/company meeting, and visited Japan more than once before.

Intention to return. For the purposes of better classification with decision trees, the variable *satisfaction* was recoded into a binary variable where 1 is highly likely and likely to return and 0 is everything else. That produced binary values that were rather equally distributed between 1 (49.1%) and 0 (50.9%). The binary response was equally distributed. The overall model's misclassification rate is 0.13, indicating an accuracy for the model of 87%.

Results-Intention to return. The odds ratio of tourists having an intention to return is higher by

- 3.9 if the tourists experienced Japanese food, want to experience Japanese sightseeing (e.g., nature, scenery) in the future, paid no higher than \$1,670 for airfare, visited Japan for the first time, and came through airports such as Narita, New Chitose (Sapporo), or Fukuoka;
- 3.9 if the tourists experienced a festival/event, sightseeing (nature/scenery), Japanese food, paid no higher than \$1,670 for airfare, and visited Japan several times;

- 1.94 if tourists experienced Japanese food, want to experience sightseeing and/or Japanese hot springs, and came with family, spouse, or friends; and
- 1.49 if tourists want to experience sightseeing in the future, experienced Japanese food, and paid no higher than \$1,670 for airfare.

Discussion

Various studies (Pizam, Neumann, & Reichel, 1978; Buhalis, 2000; Weiermair, 2000; Kozak & Rimmington, 2000; Yoon & Uysal, 2005; Chen & Tsai, 2007; Liu, Sigauw, & Enz, 2008; Chen & Chen, 2010; Lee, Scott, & Kyle 2012; Ozdemir & Simsek, 2015) have acknowledged the significance of destination image attributes and their impact on tourist satisfaction and behavioral intentions. Even studies using big data sets (Yoon & Uysal, 2005; Chen & Tsai, 2007; Chi & Qu, 2008; Alegre & Garau, 2011; Lee, Scott, & Kyle 2012; B., Lee, C., Lee, & J., Lee, 2014; Rameseo-Munhurrun, Seebaluck, & Naidoo, 2015; Ozdemir and Simsek 2015; Bajs, 2015) have confirmed and emphasized the importance of Japanese food, shopping, and transportation or information about transportation to tourists' satisfaction and intention to return to Japan as a destination. Contrary to the popular belief that the Internet is the main source of information (Buhalis & Law, 2008; Litvin, Goldsmith, & Pan, 2008), Lonely Planet books were a major source of information prior to visits to Japan. However, online information was very important while in Japan for tourists to return in the future. Place of arrival (airports), prior visits to Japan, country of residency, and flight cost were among the top variables selected by the DT, perhaps due to convenience and centralization of attractions and/or businesses around airport areas. For example, Tokyo's main international airport is Narita (Tokyo is the capital and a major business and attraction center), Osaka's airport is Kansai (Osaka is a major trade center), and New Chitose is Sapporo's airport (Sapporo is the northern

island capital). An interesting point is that credit cards as a method of payment rather than cash were more important to tourists. It was also notable that satisfaction reflects similar variables with several important differences. Preference for the hotel and quality of accommodations as well as two main destinations visited become important in the model (Joppe, Martin, & Waalen, 2001).

Thus, variables show an interesting variance between satisfaction and intention to return. Most of the important variables in both models are related to convenience and food, such as Japanese food, shopping, transportation, flight cost, etc. However, the models differ in two ways: for intention to return, the source of information and desire to experience new things are important; however, for satisfaction, accommodations and destinations within Japan are of importance.

Results of decision trees indicate that there are two distinct groups, namely, Asian and non-Asian tourists, who have different preferences related to a high level of satisfaction. The main theme for non-Asian tourists is experiencing Japanese food, shopping at supermarkets, staying at Western-style hotels, staying less than eight days, and reasonable airfare costs. This findings support the results of previous research (Joppe et al., 2001; Alegre and Garau, 2011; Bajs, 2015). For Asian tourists, higher satisfaction can be achieved by those who experience Japanese food, stay at a Japanese-style inn, come mainly for an event such as a conference, for incentive travel, or to study, and by those who have visited Japan more than once.

On the other hand, for future intent to return to Japan, nationality plays no role in the decision, but whether the visitors are more family-oriented/non-business and whether they are first-time visitors have effects on their intention. The main motivation for a visitor's future return is not driven by experiences they had during their visit but rather by experiences they want to

have when they return, such as Japanese hot springs or immersing themselves in the beauty of nature. Furthermore, experiencing Japanese food appears to remain a main attraction across all segments as a common denominator to attract all different groupings.

The decision tree analyses revealed the existence of intriguing segments irrespective of nationality, gender, total expenditure during current visit, or even the purpose of the visits, such as a core repeater grouping of those who “experienced Japanese food, want to next time experience Japanese nature/scenery sightseeing, paid no higher than \$1,670 for airfare, visited Japan for the first time, and came through airports such as Narita, New Chitose (Sapporo), or Fukuoka,” a grouping whose likelihood of returning to Japan is almost four times (3.9) higher than that of average inbound visitors.

Nationality plays an important role in the satisfaction of tourists and again separates them into two distinct groups of Asian and non-Asian. An interesting point about satisfaction is that tourists from non-Asian countries have higher odds of being satisfied than tourists from Asian countries. This could potentially be explained by the desire of non-Asians to visit a destination with a culture very different from their homeland. *On the other hand, Tran and Ralston (2006) proposed a model pertaining to tourists’ unconscious needs and their preferences in tourism. For example, they found a connection between achievement motivation and preference for adventure in American tourists. Considering that the largest non-Asian group of tourists in this study is from the United States, the authors speculate that Japan falls into both categories as a very different culture and an adventurous destination (a unique country with a different culture, different food, and a different language that is a long distance away from home). Also, for Asian countries, airfare is no longer an important variable, which could probably be explained by closer proximity. For example, the likelihood of satisfaction is a little more than double (2.32) for*

“tourists from non-Asian countries who had experience with Japanese food, paid no higher than \$1,500 for airfare, purchased Japanese fruits, and shopped at a supermarket.” Conversely, the odds increase by a little more than half (1.64) “if they are from a neighboring Asian country (Korea, China, Taiwan, Hong Kong, or Thailand), stayed at a Japanese-style inn, experienced Japanese food, came for tourism/leisure, incentive travel, study, or an international conference, and came through one of two main airports (Narita/Haneda).”

Managerial Implications

Data mining is a data-driven technique and will analyze data without introducing any major subjectivity of data analysts who may have preferred approaches or agendas associated with their past research streams. Therefore, by its structure, it would not lead to the verification of a specific existing theory unless the researchers are dealing with multiple data sets and see commonalities across them.

Data mining, however, presents unique managerial implications in that the resulting analysis can more effectively identify certain combinations of profiles and characteristics of visitors without relying on the study design or the subjective judgment of the researchers. In other words, data mining results can present more compelling responses to questions of marketing return on investment (ROI) on expenditures by government marketing and destination marketing organizations (DMOs) by identifying the specific grouping of potential visitors who are more likely to come back as repeaters than any other combinations of groupings—based purely on the analysis of the objective big data in question. Odds ratios are rather easy to interpret for non-academic practitioners, and with an understanding of groupings with higher odds ratios, DMOs may put a strategy in place to aim at groups whose odds ratios are higher than

certain discretionary benchmarks. For example, a DMO's strategy to market to groups with an odds ratio higher than 1.5 means the DMO is aiming at groups who have at least a 50% higher likelihood (of being satisfied with the destination or of coming back to the destination), with an expectation that the same marketing expenditures would work more efficiently than targeting the average at large. Data mining results might not corroborate prior-knowledge, prior-beliefs, or myths, as the structure would not pay any attention to those but rather would purely seek winning combinations of groupings by mathematical computations in an objective manner. As governments and DMOs can access big data on tourists, data mining techniques would open a new chapter for their quantitative data analysis to aid managerial decisions for better allocation of their limited resources into segments more likely to be enticed back to their destination than the market average.

Study Limitations and Future Research

This study has several structural limitations, and we believe that acknowledging those limitations may lead to more viable future research in the field of quantitative destination marketing. First, our research is based on the data of visitors who came to Japan, representing a small fraction of all the travelers who decided not to visit Japan and to whom we cannot extrapolate our findings. Second, our research data was collected at one period of the study year of 2010, after which Japan saw a huge drop in visitors due to Great East Japan Earthquake and simultaneous radiation leaks from nuclear power plants in Fukushima in 2011. The total number of inbound visitors to Japan reached the national goal of 10 million in 2013, overcoming the negative impacts on inbound visitors in the two years prior. We do not have any evidence to support whether our findings based on 2010 data have temporal stability with later data. This

indicates that updated research on the data may generate an interesting answer to the issue of temporal stability of the behavior of inbound visitors.

Finally, we did not have any involvement with the data collection processes nor the design of the survey; thus, we have dealt with secondary data collected by professionals. The lack of direct data collection experience with the dataset may prevent us from having certain insights which may be useful in evaluating the dataset, or it may have just saved us from any sampling errors associated with data input. Future research may be done with a survey exclusively on satisfaction and repeat intents, without piggybacking with the visitors' expenditure survey, the length of which may be a partial reason for relatively higher numbers of incompleteness of the survey.

References

- Ahmed, Z. U. 1991. The influence of the components of a state's tourist image on product positioning strategy. *Tourism management*, 12(4): 331-340.
- Alegre, J., & Garau, J. 2011. The factor structure of tourist satisfaction at sun and sand destinations. *Journal of Travel Research*, 50(1): 78-86.
- Apté, C., & Weiss, S. 1997. "Data mining with decision trees and decision rules." *Future Generation Computer Systems*. 13(2): 197-210.
- Assaker, G., & Hallak, R. 2013. Moderating effects of tourists' novelty-seeking tendencies on destination image, visitor satisfaction, and short-and long-term revisit intentions. *Journal of Travel Research*, 0047287513478497.
- Assaker, G., Vinzi, V. E., & O'Connor, P. 2011. "Examining the effect of novelty seeking, satisfaction, and destination image on tourists' return pattern: A two factor, non-linear latent growth model." *Tourism Management*. 32(4): 890-901.
- Athiyaman, A. 1997. Knowledge development in tourism: tourism demand research. *Tourism Management*. 18(4): 221-228.
- Bajs, I. P. 2015. Tourist perceived value, relationship to satisfaction, and behavioral intentions the example of the Croatian tourist destination Dubrovnik. *Journal of Travel Research*, 54(1): 122-134.
- Baloglu, S., & McCleary, K. W. 1999. "US international pleasure travelers' images of four Mediterranean destinations: A comparison of visitors and nonvisitors." *Journal of Travel Research*. 38(2): 144-152.
- Berry, M. J., & Linoff, G. 2000. *Mastering data mining*. 1st ed. (Canada: Wiley Computer Publishing).
- Bland, J. M., & Altman, D. G. 2000. "The odds ratio." *The Bmj*. 320(7247). 1468.
<http://www.bmj.com/content/320/7247/1468.1.short>
- Brown, T. J., Barry, T. E., Dacin, P. A., & Gunst, R. F. 2005. "Spreading the word: Investigating antecedents of consumers' positive word-of-mouth intentions and behaviors in a retailing context." *Journal of the Academy of Marketing Science*. 33(2): 123-138.
- Buhalis, D. 1997 "Information technology as a strategic tool for economic, social, cultural and environmental benefits enhancement of tourism at destination regions." *Progress in tourism and hospitality research* 3 (1): 71-93.
- Buhalis, D. 1998. "Strategic use of information technologies in the tourism industry." *Tourism Management*. 19(5): 409-421.

- Buhalis, D. 1999. "Tourism on the Greek Islands: Issues of peripherality, competitiveness and development." *International Journal of Tourism Research*. 1(5): 341-358.
- Buhalis, D. 2000. "Marketing the competitive destination of the future." *Tourism Management*. 21(1): 97-116.
- Buhalis, D., & Law, R. 2008. Progress in information technology and tourism management: 20 years on and 10 years after the Internet -The state of eTourism research. *Tourism management*, 29(4): 609-623.
- Buhalis, D., & Law, R. 2008. "Progress in information technology and tourism management: 20 years on and 10 years after the Internet -The state of eTourism research." *Tourism Management*. 29(4): 609-623.
- Chang, K., Chen, M., Kuo, N., Hsu, C., & Cheng, Y. 2016. Applying data mining methods to tourist loyalty intentions in the international tourist hotel sector. *Anatolia: An International Journal of Tourism & Hospitality Research*, 27(2): 271-274.
- Chen, C. F., & Tsai, D. 2007. How destination image and evaluative factors affect behavioral intentions? *Tourism management*, 28(4): 1115-1122.
- Chi, C. G. Q., & Qu, H. 2008. Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tourism management*, 29(4): 624-636.
- Chon, K. S., & Olsen, M. D. 1991. "Functional and symbolic congruity approaches to consumer satisfaction/dissatisfaction in tourism." *Journal of the International Academy of Hospitality Research* (3), 2-22.
- David Walton (March 28, 2014) Big Data raises big legal issues. Inside Counsel. Retrieved from: <http://www.insidecounsel.com/2014/03/28/big-data-raises-big-legal-issues>
- De Reyck, B., Degraeve, Z., & Vandenborre, R. 2008. "Project options valuation with net present value and decision tree analysis." *European Journal of Operational Research*, 184(1): 341-355.
- Duncan, R. 1980. "What is the right organization structure? Decision tree analysis provides the answer." *Organizational Dynamics*, 7(3): 59-80.
- Dwyer, L., Pham, T., Forsyth, P., & Spurr, R. 2014. "Destination Marketing of Australia Return on Investment." *Journal of Travel Research*, 53(3): 281-295.
- Fakeye, P. C., & Crompton, J. L. 1991. "Image differences between prospective, first-time, and repeat visitors to the Lower Rio Grande Valley." *Journal of travel research*, 30(2): 10-16.
- Fayyad, U. M., Piatetsky-Shapiro G., Smyth P., and Uthurusamy R., (1996). "Advances in knowledge discovery and data mining." *AI Magazine* 17(3): 37- 54

- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. 1992. "Knowledge discovery in databases: An overview." *AI Magazine*, 13(3): 57.
- Fuchs, G., & Reichel, A. 2011. "An exploratory inquiry into destination risk perceptions and risk reduction strategies of first time vs. repeat visitors to a highly volatile destination." *Tourism Management*, 32(2): 266-276.
- Gallarza, M. G., & Saura, I. G. 2006. "Value dimensions, perceived value, satisfaction and loyalty: an investigation of university students' travel behavior." *Tourism Management*, 27(3): 437-452.
- Goh, C., Law, R., & Mok, H. M. 2008. "Analyzing and forecasting tourism demand: A rough sets approach." *Journal of Travel Research*, 46(3): 327-338.
- Hall, C. M. and Hall, M., *Tourism planning: policies, processes and relationships*. Upper Saddle River, NJ: Pearson Education, 2000.
- Hallowell, R. 1996. "The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study." *International Journal of Service Industry Management*, 7(4): 27-42.
- Hernández Lobato, L., Solis-Radilla, M. M., Moliner-Tena, M. A., & Sánchez-García, J. 2006. "Tourism destination image, satisfaction and loyalty: a study in Ixtapa-Zihuatanejo, Mexico." *Tourism Geographies*, 8(4): 343-358.
- Hosking, J. R., Pednault, E. P., & Sudan, M. 1997. "A statistical perspective on data mining." *Future Generation Computer Systems*, 13(2): 117-134.
- Japan National Tourism Organization 2012. "Foreign Visitors and Japanese Departures." Report prepared by Japan National Tourism Organization. (Accessed on November 2014).
- Japan Tourism Research and Consulting Co. (2016) Historical Statistics-Visitors to Japan from Overseas. Report prepared by National Tourism Organization (accessed on July, 2016) Retrieved from: <http://www.tourism.jp/en/statistics/inbound/>
- Joppe, M., Martin, D. W., & Waalen, J. (2001). Toronto's image as a destination: A comparative importance-satisfaction analysis by origin of visitor. *Journal of travel research*, 39(3), 252-260.
- Pagliery J., (May 8, 2015) OMG: 2.1 million people still use AOL dial-up. CNN Money. Retrieved from: <http://money.cnn.com/2015/05/08/technology/aol-dial-up/>
- Kim, S. S., Timothy, D. J., & Hwang, J. (2011). Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method. *Tourism Management*, 32(3), 544-554.
- Kozak, M. 2001. "Comparative assessment of tourist satisfaction with destinations across two nationalities." *Tourism Management*. 22(4): 391-401.

- Kozak, M. 2001. "Repeaters' behavior at two distinct destinations." *Annals of Tourism Research*, 28(3): 784-807.
- Kozak, M. 2002. Comparative analysis of tourist motivations by nationality and destinations. *Tourism Management*. 23(3): 221-232.
- Kozak, M. 2003. "Measuring tourist satisfaction with multiple destination attributes." *Tourism Analysis*. 7(3-4): 229-240.
- Kozak, M., & Rimmington, M. 2000. "Tourist satisfaction with Mallorca, Spain, as an off-season holiday destination." *Journal of Travel Research*, 38(3): 260-269.
- Laesser, C., & Dolnicar, S. 2012. "Impulse purchasing in tourism—learnings from a study in a matured market." *Anatolia*, 23(2): 268-286.
- Lee, J. J., Kyle, G., & Scott, D. (2012). The mediating effect of place attachment on the relationship between festival satisfaction and loyalty to the festival hosting destination. *Journal of Travel Research*, 51: 754-767
- Lee, B., Lee, C. K., & Lee, J. (2014). Dynamic nature of destination image and influence of tourist overall satisfaction on image modification. *Journal of Travel Research*, 53(2), 239-251.
- Leiper, N. 1995. *Tourism Management* (Vol. 455). French Forest: Pearson Education Australia.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, 29(3), 458-468.
- Liu, Z., Siguaw, J. A., & Enz, C. A. (2008). Using tourist travel habits and preferences to assess strategic destination positioning the case of Costa Rica. *Cornell Hospitality Quarterly*, 49(3): 258-281
- Lupton, R. A. 1997. "Customer portfolio development: Modeling destination adopters, inactives, and rejecters." *Journal of Travel Research*, 36(1): 35-43.
- Maimon, O., and Rokach, L. 2005. *Decomposition methodology for knowledge discovery and data mining* (pp. 981-1003). Springer US.
- Mannila, H. 2000. "Theoretical frameworks for data mining. *ACM SIGKDD*". *Explorations Newsletter*, 1(2): 30-32.
- Maxham, J. G. 2001. "Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions." *Journal of Business Research*. 54(1): 11-24.
- Min, H., Min, H., & Emam, A. (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management*, 14(6), 274-285

- Miner, G., Nisbet, R., & Elder IV, J. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press.
- Mishler, A. L. 1965. "Personal Contact in International Exchange." In *International Behavior, A Social-Psychological Analysis*, edited by Herbert C Kelman. New York: Holt, Rinehart & Winston.
- Nisbet, R., Elder IV, J., & Miner, G. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press Burlington, MA: Elsevier.
- Okamura, K., & Fukushige, M. 2010. "Differences in travel objectives between first-time and repeat tourists: An empirical analysis for the Kansai area in Japan." *International Journal of Tourism Research*, 12(6): 647-664.
- Özdemir, G., & Şimşek, Ö. F. (2015). The Antecedents of Complex Destination Image. *Procedia-Social and Behavioral Sciences*, 175, 503-510.
- Park, Y. A., & Gretzel, U. 2007. "Success factors for destination marketing web sites: A qualitative meta-analysis." *Journal of Travel Research*, 46(1), 46-63.
- Pearce, D. G. 2014. "Toward an integrative conceptual framework of destinations." *Journal of Travel Research*, 53(2): 141-153.
- Phillips, W. J., & Back, K. J. 2011. "Conspicuous consumption applied to tourism destination." *Journal of Travel & Tourism Marketing*, 28(6): 583-597.
- Pike, S. 2012. *Destination Marketing*. 1st ed. Burlington, MA USA: Elsevier
- Pizam, A., Neumann, Y., & Reichel, A. 1978. "Dimensions of tourist satisfaction with a destination area." *Annals of Tourism Research*, 5(3): 314-322.
- Pool, I. D. 1965. "Effects of Cross-National Contact on Nation and International Images" In *International Behavior: A Social Psychological Analysis*, edited by Herbert C Kelman. New York: Holt, Rinehart & Winston.
- Pyo, S., Uysal, M., & Chang, H. 2002. "Knowledge discovery in database for tourist destinations." *Journal of Travel Research*. 40(4): 374-384.
- Qu, Y., Adam, B. L., Yasui, Y., Ward, M. D., Cazares, L. H., Schellhammer, P. F., & Wright, G. L. 2002. "Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients." *Clinical Chemistry*, 48(10): 1835-1843.
- Ranaweera, C., & Prabhu, J. 2003. "On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth." *Journal of Targeting, Measurement and Analysis for Marketing*, 12(1): 82-90.

- Ramseook-Munhurrun, P., Seebaluck, V. N., & Naidoo, P. 2015. Examining the structural relationships of destination image, perceived value, tourist satisfaction and loyalty: case of Mauritius. *Procedia-Social and Behavioral Sciences*, 175: 252-259.
- Reinartz, W., & Kumar, V. 2002. "The Mismanagement of Customer Loyalty." *Harvard Business Review*, 80(7): 86-94.
<https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=6899224&site=eds-live&scope=site>
- Maimon, O., & Rokach, L. (2005). Decomposition methodology for knowledge discovery and data mining. In *Data Mining and Knowledge Discovery Handbook* (pp. 981-1003). Springer US.
- Ryan, C. 1991. "Tourism and marketing—a symbiotic relationship?" *Tourism Management*, 12(2): 101-111.
- Sheldon, P. 1994. "Information technology and computer systems." *Tourism Marketing and Management Handbook*, 2.
- Som, A. P. M., & Badarneh, M. B. 2011. "Tourist satisfaction and repeat visitation; toward a new comprehensive model." *International Journal of Human and Social Sciences*, 6(1): 38-45.
- Tran, X., & Ralston, L. (2006). Tourist preferences influence of unconscious needs. *Annals of Tourism Research*, 33(2), 424-441.
- UNWTO Policy and practice for global tourism UNWTO, Madrid (2011)
- Uzama, A. 2009. "Marketing Japan's travel and tourism industry to international tourists." *International Journal of Contemporary Hospitality Management*, 21(3): 356-365.
- Valle, P. O. D., Silva, J. A., Mendes, J., & Guerreiro, M. 2006. "Tourist satisfaction and destination loyalty intention: a structural and categorical analysis." *International Journal of Business Science and Applied Management*, 1(1): 25-44.
- Wang, Y., & Pizam, A. 2011. *Destination marketing and management: theories and applications*. 1st ed. Oxfordshire, UK: CABI.
- Weiermair, K. 2000. "Tourists' perceptions towards and satisfaction with service quality in the cross-cultural service encounter: implications for hospitality and tourism management." *Managing Service Quality: An International Journal*, 10(6): 397-409.
- Wicker, P., & Breuer, C. 2013. "Exploring the critical determinants of organizational problems using data mining techniques: evidence from non-profit sports clubs in Germany." *Managing Leisure*, 18(2): 118-134.
- Woodside, A. G., & Lysonski, S. 1989. "A general model of traveler destination choice." *Journal of Travel Research*, 27(4): 8-14.

- Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism management*,26(1), 45-56.
- Yuksel, A., Yuksel, F., & Bilim, Y. (2010). Destination attachment: Effects on customer satisfaction and cognitive, affective and conative loyalty. *Tourism Management*, 31(2), 274-284.

Tables and Figures

Table 1 Analysis of Relevant Past Research

Author(s)	Purpose(s)	Objects of Observation	Findings
Yoon and Uysal 2005	to understand tourist motivations based on the push and pull motivators, satisfaction and destination loyalty	Northern Cyprus	Tourist destination loyalty is effected by satisfaction and experiences
Chen and Tsai 2007	to construct an integrated model of tourist consumption process and to examine relationships between destination image, evaluative factors and behavioral intentions.	Kengtin region in southern Taiwan	Destination image has most important effect on behavioral intentions directly and indirectly.
Chi and Qu 2008	to offer an integrate approach for understanding destination and examine an empirical evidence on the casual relationships among destination image, tourist satisfaction and loyalty	Arkanzas, Eureka Springs, USA	Statistically significant relationship among destination image, attribute satisfaction, overall satisfaction and destination loyalty. Full mediation role of overall satisfaction between destination image and loyalty
Alegre and Garau 2011	to identify key drivers of sun and sand product on tourist satisfaction using penalty-reward analysis	Palma airport, Spain	Penalty-reward method supported an asymmetrical relationship between satisfaction with attributes and overall satisfaction.
Lee, Lee, and Lee 2014	to investigate dynamic nature of destination image and role of satisfaction in modifying it	Seoul, Korea	There is a significant difference in the destination image between pre and post trip
Rameseoo-Munhurrun, Seebaluck and Naidoo 2015	to develop a conceptual model for destination image	Mauririus, East Africa	Destination image and perceived value are direct determinants of satisfaction

Ozdemir and Simsek 2015	to examine a complex destination images by evaluating interconnecting relationships between destination image, perceived price, quality, value and overall satisfaction.	Izmir, Turkey	Tourist induced destination image changes only based on the experiences he/she has
Lee, Scott and Kyle 2012	to explore factors that drive festival visitors loyalty to host destinations	Pasadena, Texas USA	Place attachment plays a mediating role in the relationship between festival satisfaction and destination royalty
Bajs 2015	to define model of tourist perceived value that was affected primarily by destination appearance and emotional experience	Dubrovnik, Croatia	Perceived value has a significant effect on future intentions and satisfaction

Table 2 Variables in order of importance for satisfaction

Variable Name	Importance
Experienced Japanese Food	1.000
Shopping	0.927
Availability of information about transportation	0.295
Lonely Planet as source of information prior to visit to Japan prior to the visit	0.285
Country of residence	0.209
Nationality	0.202
Airport of entry	0.124
Main destination visited in Japan	0.117
Main purpose of the visit to Japan	0.110
Secondary destination visited in Japan	0.091
Expenditure at the place of stay (hotel, etc.,)	0.090
Prior visit to Japan	0.090
Level of expectation for business trip	0.083
Cosmetics and pharmacy expenditure	0.074
Credit card as a method of payment in Japan	0.073

Table 3 Variable importance for intention to return

Variable	Importance
Experienced Japanese Food	1.000
Shopping	0.930
Transportation	0.301
Lonely Planet as a major source of information about Japan prior to visit	0.293
Which airport did you land in Japan	0.134
How many time have you visited Japan including this visit	0.124
Main area (destination) in Japan visited	0.094
Internet as a main helpful source in obtaining information while in Japan	0.090
Desire to experience nature/scenery sightseeing next visit	0.081
Flight cost	0.077
Country of residency	0.075
Want to walk around downtown in the future	0.069
Catering cost	0.064
Cosmetics and pharmacy expenditure	0.062
Nationality	0.060

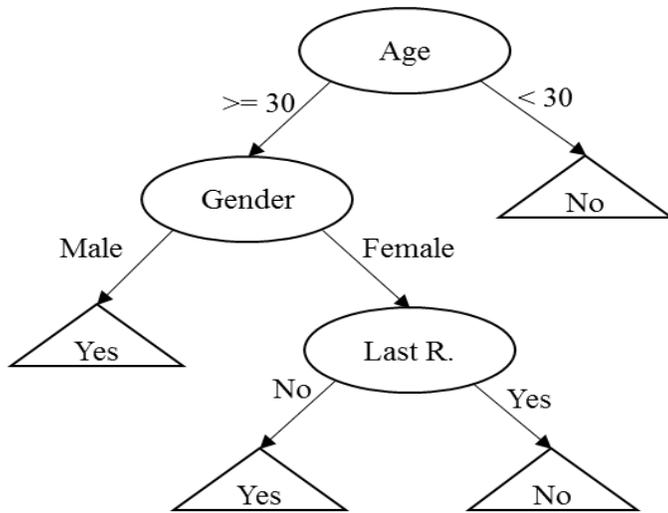


Figure 1 Decision tree sample adapted from (Rokach & Maimon, 2010).

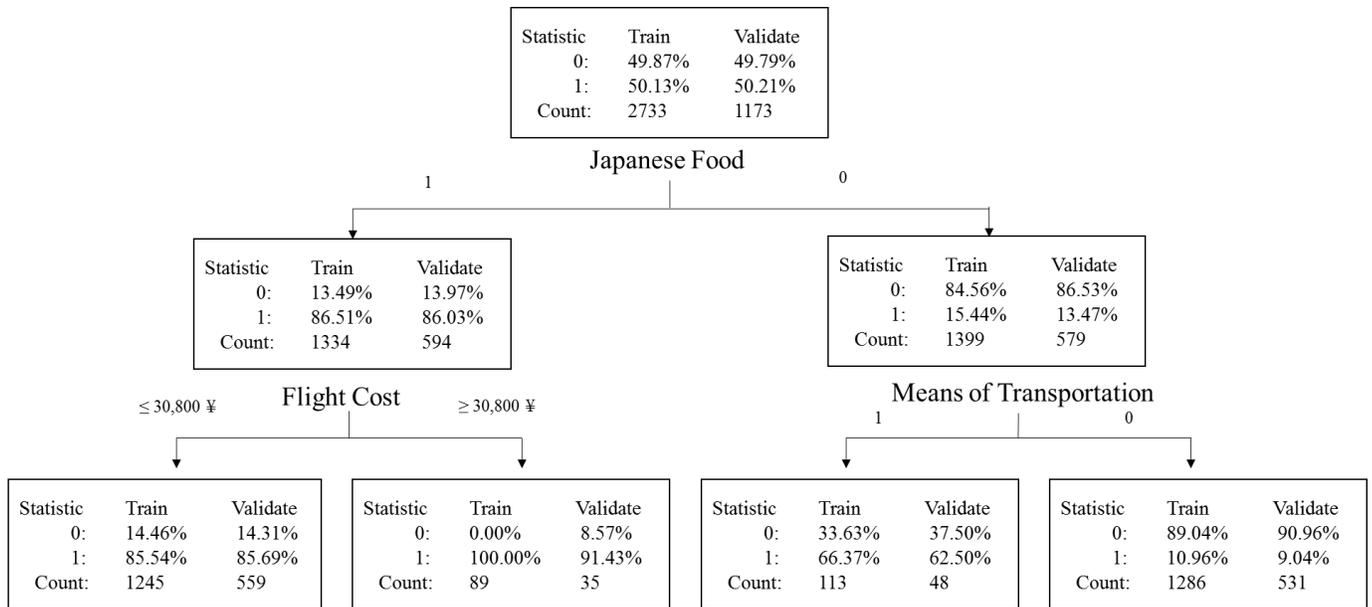


Figure 3 Excerpt of decision tree for Intention to return