Corporate Management Review
Vol. 45 No. 2, 2025
pp. 37-100

A sentiment analysis-based method for target market decision-making using big data from travel communities
一個基於旅遊社群大數據情感分析的目標市場決策方法

Hsin-Hui Lin[1]
Department of Distribution Management, National Taichung University of Science and Technology
Wan-Chen Lee
Department of Distribution Management, National Taichung University of Science and Technology

Abstract: This study aims to develop a method for making target market decisions based on sentiment analyses of travel communities' big data. This study first adopts a qualitative sentiment analysis method involving the use of Python Web Crawler technology to collect customer comments on a hotel booking website, Booking.com, and a travel itinerary website, Klook. The mean score and importance of each word in the comments are calculated for the qualitative sentiment analysis. Later, based on the sentiment analysis results, this study applies the quantitative decision tree method, and a questionnaire is designed. The 305 valid responses obtained are analyzed using the decision tree method in SPSS Modeler to segment the target market. Based on the sentiment analysis and the decision tree analysis results, this study provides marketing suggestions to travel services platforms and providers regarding the various target market segments. This study is a pioneering effort to develop a method for making target market decisions based on sentiment analyses of travel communities' big data.

Keywords: Target market decision making, travel community, sentiment analysis, decision tree.

[1] Corresponding author: Hsin-Hui Lin, Department of Distribution Management National Taichung University of Science and Technology. E-mail: brenda@nutc.edu.tw

DOI: 10.53106/102873102025124502002

摘要：本研究旨在基於旅遊社群大數據的情感分析，開發一種目標市場決策方法。首先，本研究採用定性情感分析方法，利用 Python 網頁爬蟲技術收集酒店預訂網站 Booking.com 和旅遊行程網站 Klook 的顧客評論，並計算評論中每個詞彙的平均分數和重要性，以進行定性情感分析。接著，根據情感分析結果，本研究進一步應用量化決策樹方法，設計問卷並收集有效回應共 305 份，隨後利用 SPSS Modeler 中的決策樹方法對目標市場進行細分。根據情感分析與決策樹分析結果，本研究為旅遊服務平台與供應商提供針對不同目標市場細分的行銷建議。本研究是基於旅遊社群大數據的情感分析來制定目標市場決策的先導性研究。

關鍵詞：目標市場決策、旅遊社群、情感分析、決策樹。

1. Introduction

Continuous economic development and the gradual rise in living standards have driven many people to pursue spiritual enjoyment and make travel an important part of their lives. Over the past decade, the number of people traveling abroad has increased year by year, and self-guided tours have become the main choice for overseas tours. Due to booming development in terms of online travel agents, travel cost comparison websites, destination travel platforms, and mobile applications, abundant and diversified travel information has become easily accessible, significantly lowering the difficulties of planning and booking self-guided tours. As a result, this type of travel method is increasingly popular with domestic people. New-generation consumers are highly reliant on online services due to the convenience they offer with respect to flight services, rental services, and other services. Encounters between the tourist and the service provider constitute an experience (Sorensen and Jensen, 2015): this includes the tour itself and any post-tour events (Shaw and Williams, 2009). For self-guided tourists, a tour experience is likely to include their feeling about all the activities involved in the tour, including online booking of flight ticket, hotels, and restaurants (Sorensen and Jensen, 2015). Compared to group tourists, this type of tourist is more likely to leave comments on the Internet and be affected by other people's comments when making a traveling decision.

Comments and posts on social media and travel platforms allow service providers to extract a large amount of structured and unstructured data, such as tourists' opinions, preferences, needs, and attitudes (Morabito, 2015). Moreover, each stage of the journey provides copious data to review. Through the integration of travel data resources and application of big data technologies, service providers can provide customized services to satisfy the unmet needs of customers, which is a critical step to value creation. Further, service providers can utilize the accumulated data to analyze markets and support their management teams. As such, big data has become an important strategic asset for service providers.

The development of big data will drive upgrades in the tourism industry, as greater emphasis is placed on analyzing customers' emotions and experiences with tour products. The tourism market is large and encompasses a wide variety of products. It is hard for service providers to provide a customized product to each tourist. Target market segmentation is a powerful marketing tool that allows service providers to capture the needs of each customer segment (McCleary, 1995) so that they are able to provide products and services that are favored and valued by their target market (Lee et al., 2004).

Applications of big data technologies in tour planning and management are complicated and require field knowledge (Chen et al., 2012). Extant research on the applications of big data covers diverse fields including health care, astronomy, social networking, earth science, and crime detection (Hashem et al., 2015). Also, some recent studies focus on using the sentiment analysis of big data in different sectors (e.g., Jena, 2020; Basiri et al., 2021; Meenaa et al., 2023). However, there has been scant discussion pertaining to electronic travel platforms. Our methodology combines both qualitative and quantitative methods. In the qualitative aspect, we conduct a text and content analysis of comments collected from various travel websites and then summarize, organize, and categorize both the comments and the sentiments expressed in the comments. On the quantitative side, we employ demographic, psychological, and behavioral variables for market segmentation, where the psychological variables serve as the foundation for segmenting tourists via factor analysis and cluster analysis. Finally, we apply the decision tree method to segment the target market and provide marketing suggestions to tour service providers. The objectives of this study are to contribute to improvements in and innovation of tour products in two ways: (1) apply sentiment analysis of comments made by tourists, establish a comment categorization framework, and develop marketing strategies; and (2) use a decision tree as the method and market segmentation variables as the basis to segment the market, predict the needs of the target market segments, and provide effective marketing strategies for each segment to tour service providers.

2. Literature review

2.1 The characteristics and patterns of travel e-commerce websites

Travel e-commerce is a travel information-based business system that enables buying and selling of travel related products and services via electronic mechanisms. In the tourism industry, e-commerce is recognized as a key factor associated with business opportunities, reduced operating costs, and enhanced work efficiency. Through e-commerce, the tourism industry can extend their customer reach to the global market, thereby creating a competitive advantage. This is because e-commerce allows them to integrate various supply chain activities and provide customized and enriched travel experiences to customers. At present, almost all business and marketing operations in the tourism industry are done online. Further, customers often use online platforms to search for information about vacation or travel destinations and purchase desirable services. Online platforms are therefore an indispensable competitive factor in the tourism industry.

The tourism industry has taken the advantage of e-commerce to make a structural industry change and generate new business opportunities in response (Seigel, 2004). Many organizations work with travel websites to provide services to their customers. As a result, various e-commerce models have emerged, with four predominant classifications based on the form of transaction. First, **B2C**: tour service providers sell products directly to customers. This model avoids information asymmetries based on distance and is one of the most extensively used in the industry. Second, **B2B**: the tourism industry is integrated, such that they have agency, transaction, and cooperation relationships with firms in the food, accommodation, transportation, travel, shopping, and entertainment industries. There is a still significant room for development in terms of this model. Third, **B2B2C**: this model combines B2B and B2C for a complete product or service transaction. Lastly, **C2C**: in this model, tourists can exchange opinions with each other through online platforms (e.g., forums, Facebook, and blogs) to make travel plans and get help with booking hotels, flight tickets or using a website (Scott's Directories, 2002). By operation type, they can be classified into one of four types. The first is sale of itinerary products: these include traditional travel agents, online travel agents, inbound travel agents, and destination travel agents. Colatour and Klook are examples—their major business area consists of selling self-developed and tour packages. The second is price search and comparison: these providers collect real-time price data on flight tickets, hotel accommodations, and tour packages offered by major travel agencies. For example, the global hotel reservation website Booking.com provides real-time price data, allowing users to compare prices and directly make reservations online. The third is provision of travel tools: people most commonly use this type to search for media reports, netizens' comments, original content, or travel experts' experiences and suggestions when they are planning a trip. Pixnet and Backpackers are examples of this type. The last is Travel P2P: this is a platform for person-to-person transaction of services based on a sharing economy model that has emerged in recent years. Uber and Airbnb are based on this model (Tsai, 2016).

2.2 Tourism big data concept and analytic types

2.2.1 Tourism big data and its applications

Big data has been recognized as a new driver of the innovation of business processes (Wamba et al., 2015), because of its potential to help create new value. Big data is also a strategy that can be employed to support analyses of various types of information. It has become a trend that many organizations are adopting to acquire and assemble valuable information (Sivarajah et al., 2017). Laney (2001) introduced a 3Vs model of big data, suggesting that big data refers to high volume, high velocity, and high variety information assets. IDC iView (2015) proposed the addition of another V—value—to expand the model to a 4Vs model. Lu et al. (2014) incorporated the concept of veracity as another feature of big data, resulting in a 5Vs model. The model expanded to 7Vs after Khan, Gupta and Uddin (2014) proposed to include validity and volatility. The 7Vs are volume, variety, velocity, veracity, value, validity, and volatility.

Tourism data is being updated all the time. A considerable portion consists of users' original their shared comments, posts, and photos (Hawelka et al., 2014). Thanks to the prevalence of the Internet and the convenience of information acquisition, various types of word-of-mouth information and consumer comments are shared and searched for: both have become an important reference for consumers and service providers. For example, they can be used to estimate the number of visitors to a specific tourist spot, identify tourist hotspots in a city (García-Palomares et al., 2015), or support formulation of a marketing strategy (Marine-Roig and Anton Clavé, 2015). The growth of big data is very important for the evaluation of travel services, as it allows for insights into tourists' preferences, thoughts, values, and behaviors (Oussous et al., 2018), leading to predictions about future likely destinations. In turn, providers can instantly recommend hotels, restaurants, and activities based on their preferences.

In addition, researchers can employ various analysis methods to convert raw data into operable marketing knowledge as illustrated by the marketing mix framework for big data analytics (Fan et al., 2015). In the "people" perspective of this framework, customer differentiation and customer profile analysis are critical to effective marketing. It is important to identify similar preferences in a specific customer group and respond to specific marketing instructions. The customer differentiation procedure facilitates the identification of different customers with similar interests. In the "product" perspective, product and reputation management can be used to reduce product weaknesses identified through the survey data. In the "promotion" perspective, in order to enhance customers' product awareness and attract potential customers, the recommendation system has been extensively applied in the e-commerce context (Chen et al., 2012). User comment-based collaborative filtering and content association mining are commonly adopted in this system. In the "price" perspective, pricing strategy analysis and competitor analysis can be applied to find the suggested pricing strategy for a specific context. Traditionally, empirical studies of pricing strategies have relied on survey data and the regression method. For example, Peter and Thomas (1999) explored the determining factors of pricing strategies through a mail survey. In the "price" perspective, location-based advertising and dynamic analysis are important dimensions of marketing analysis. Research on location-based advertising is focused mainly on the effect of location in a marketing strategy. For example, a study collected customer data from a survey to analyze the regional strategies used by wineries in brand promotion.

2.2.2 Social media and types of big data analytics

According to Kaplan and Haenlein (2010), social media is built on two elements, namely Web 2.0 and user-generated content. Defining social media as a social website that provides online services, Ellison (2007) identified three characteristics of social media: it allows for the creation of public or semi-public data by individual users; it permits users to establish a network with other users; and it lets users to view the activities associated with other users and make them publicly accessible on the network. People can share their opinions, good or bad, on topics, products, and services through the Internet and social media. Social media platforms support numerous data formats, including text, graphic, video, audio, and geolocation. From customer information and big data, companies can uncover hidden messages about their customers to take preventive action, enhance customer satisfaction, and discover new opportunities. They can also predict their tendencies and intentions and translate possibilities into opportunities to maximize the business value of each customer. The data can be classified into unstructured and structured data (Baar and Kemper, 2008).

Big data from social media use combined with advancements in computing tools have become the key to understanding human behavior. Big data is constantly being stored and processed by companies, individuals, and governments (Manovich, 2011). The most common applications of big data in social media include trend discovery, social media analysis, sentiment analysis, and opinion mining. Customer feedback collected from social media can be used to modify decisions and derive business value (Wu et al., 2014). Generally, there are 13 types of social media: blogs, business networks, collaborative projects, enterprise social networking, forums, microblogs, photo sharing, product/service reviews, social bookmarking, social gaming, social networks, video sharing, and virtual worlds (Thomas and Frank, 2015). In this study, we collect reviews posted on travel product/service platforms including a hotel reservation website, Booking.com, and a travel itinerary website, Klook, for subsequent analysis.

A large body of research is focused on analysis techniques for social media big data (Adedoyin-Olowe et al., 2014 ; Park et al., 2017; Lv et al., 2017; Kim and Hastak, 2018). There are four types including: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics.

2.3 The concept and methods of sentiment analysis

Sentiment analysis, also referred to as opinion mining, is one key area in big data research. The first step of sentiment analysis is to extract emotions, feelings, or opinions from text messages in web content (Fang and Zhan, 2015). Later, the extracted data needs to be classified as positive, negative, and neutral, which respectively suggest liking, disliking, and indifference about a certain event or product (Ohbe et al., 2019). In the tourism industry, the data for sentiment analysis is usually obtained from comments posted on travel websites and social media.

There are two approaches to sentiment analysis: the machine learning approach and lexicon-based approach. Machine learning is an area of artificial intelligence. It is based on selection of an appropriate feature set to detect emotions. However, due to the unique characteristic of social media and the extensive use of slang and colloquialisms in many posts, big data collected from social media usually cannot be directly used. Without a conversion into meaningful insights, the data cannot effectively support decision-making (Gandomi and Haider, 2015). Lexicon-based approach relies on a collection of known sentiment terms, so it can be further divided into a corpus-based approach and a dictionary-based approach. The former calls for the collection of a large number of documents followed by the use of statistical methods to observe regularities, mine the concept words, and detect the polarity of each word. It relies on a large corpus for training and learning: words can have numerous meanings, so incorrect judgments of a word's polarity are possible. The latter approach employs artificial intelligence technologies to first collect a portion of sentiment words in the given text. These sentiment words are used as seed sentiment words. Later, it adopts an online dictionary, such as HowNet, to find the synonyms and antonyms for each word and determine word polarity through comparison. We use the dictionary-based approach for sentiment analysis in this study.

Data scientists analyze text and opinions through natural language processing (NLP). In NLP, each word is considered as a basic unit. The frequency of each word or the relationship in the glossary is analyzed (Brooker et al., 2016). Word segmentation is an important technique in language processing. The most basic task of NLP is to correctly divide a string of words into multiple component words. The Chinese Knowledge and Information Processing (CKIP) Word Segmentation System developed by Academia Sinica in

Taiwan is a more fully-established system. It is adopted in this study to segment words in the collected comments. So far, numerous sentiment dictionaries have been developed. The more commonly used Chinese sentiment dictionaries include HowNet, National Taiwan University Sentiment Dictionary (NTUSD), and E-HowNet; more commonly used English sentiment dictionaries include WordNet, MPQA, AFINR, and the Bing Liu opinion dictionary.

Recent advancements in big data analytics and sentiment analysis have significantly influenced decision-making processes in the tourism industry. By analyzing user-generated content from travel communities, businesses can gain insights into consumer sentiments, enabling more informed target market decisions.

Alaei et al. (2019) conducted a comprehensive review of sentiment analysis applications in tourism, highlighting how analyzing online reviews and social media content can provide valuable insights into tourist perceptions and preferences. Their study emphasizes the importance of automated sentiment analysis in managing the vast amount of data generated online, which can inform marketing strategies and enhance customer satisfaction. Fronzetti Colladon et al. (2021) applied social network and semantic analysis to online travel forums to forecast tourism demand. By examining over 2.6 million posts from TripAdvisor forums, they identified that variables such as language complexity and the centralization of communication networks significantly contribute to predicting international airport arrivals. This approach demonstrates the potential of sentiment analysis in understanding traveler behavior and improving demand forecasting. De Fretes et al. (2024) utilized sentiment analysis on social media data to assess tourism's social and emotional impact in Indonesia. Their findings revealed that positive sentiments dominated online discussions about Indonesian tourism, indicating strong enthusiasm for the country's travel experiences. Such insights can guide stakeholders in developing marketing strategies that resonate with target audiences and address areas of concern highlighted by negative sentiments.

In summary, leveraging sentiment analysis of big data from travel communities enables tourism businesses to make data-driven decisions regarding target markets. By understanding consumer sentiments and preferences, companies can tailor their marketing strategies to enhance customer engagement and satisfaction.

2.4 Target customers' travel behavior and decision tree

2.4.1 Marketing decisions and target market segmentation

By combining target customers' information and big data, companies can extract hidden knowledge about their customers and turn it into opportunities to maximize the value of each customer, take preventive measures, improve customer satisfaction, discover new opportunities, and predict the tendencies and intentions of their customers (Marshall et al., 2015). Based on the obtained insights, companies can also adjust their product mix and optimize internal processes and production processes. With the rapid development of the Internet and the increasing prevalence of e-commerce, companies have developed numerous ways to collect diverse information from customers (Chang et al., 2013).

Marketing is a key to profit generation. As consumers become more and more knowledgeable about both products and their rights, the importance of precision marketing is more likely to be recognized. Precision marketing requires collecting data from specific customers based on their personal characteristics. With customer data and transaction records, businesses can gain insights into customer behavior and preferences to more accurately design a marketing mix and meet the needs of the target market. Market characteristic research usually segments a market based on the following variables: (1) demographic variables: gender, age, marital status, and family size; (2) geographic variables: country, location of residence, location of workplace; (3) psychological variables: lifestyle and attitude; and (4) behavioral variables: buying behavior, sales volume, number of purchases, purchase frequency, amount of spending, time of purchase, and duration of consumption.

2.4.2 Application of decision tree analysis in marketing

A decision tree is a classical data mining algorithm, which works like the human decision-making process and can handle both discrete and continuous data (Banu, 2016). It is popular in the field of feature selection (Sun and Hu, 2017). A decision tree diagram is a tree structure in which each internal node represents a test on an attribute, each branch represents the outcome of the test, each leaf node represents a class label, and the top node represents the root (Srivastava et al., 1999).

Decision trees are usually used to solve classification problems in machine learning (Quinlan, 1993). Classification analysis is a process of classifying a group of known objects according to the defined attributes, with the goal of creating a training model. The created model can be used to infer decision rules from known data, predict the type or value of the target variable, and find the decision variables (Chauhan and Thakur, 2014). Each piece of data in the dataset has multiple features, and each feature has a different impact on the prediction result. Through the use of algorithms, the amount of information of each feature after classification can be measured. The more commonly used decision tree algorithms include the Chi-square Automatic Interaction Detector (CHAID) (Hartigan, 1975), the Classification & Regression Tree (C&RT) (Breiman et al., 1984), ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), and Neural Networks (Rumelhart et al., 1993).

In the field of marketing, decision trees are mainly applied in target marketing, customer relationship management, and customer consumption pattern research. As to their application in market segmentation, Lynd (2002) mentioned that because data mining software is highly accessibility and ease of use of, the application of the decision tree method in market segmentation has become increasingly common. In this study, we leverage the decision tree method for classification to effectively identify the market segments for the travel industry, build a decision model for precision marketing, and provide appropriate strategies to explore the target market and meet the needs of target customers.

3. Method

3.1 Method 1: Sentiment analysis

3.1.1 Analysis framework

The following steps were taken to collect sentiments and comments generated by travelers before and after buying products or services on travel platforms. First, Python Web Crawler technology was utilized to collect consumers' comments on Booking.com and Klook. The collected comments were processed using the CKIP Word Segmentation System developed by Academia Sinica to divide the original text into multiple units for subsequent analysis. Later, based on the dictionary-based approach of sentiment analysis, the opinion words, degree words, and negation words in the text were identified and compared with the reference words in the E-HowNet sentiment dictionary (containing opinion words carrying a positive or negative emotion, degree words that express the intensity of emotions, and negation words that express negative feeling). For words that could not be compared or words without a clear idea, the content coding method was used. Three experts in the area of tourism management were invited to measure the reliability of the coding results. Finally, the product of the score for the opinion word and the weight of the degree word was calculated and used as the score for the comment. The analysis framework of Method 1 is as illustrated in Figure 1.

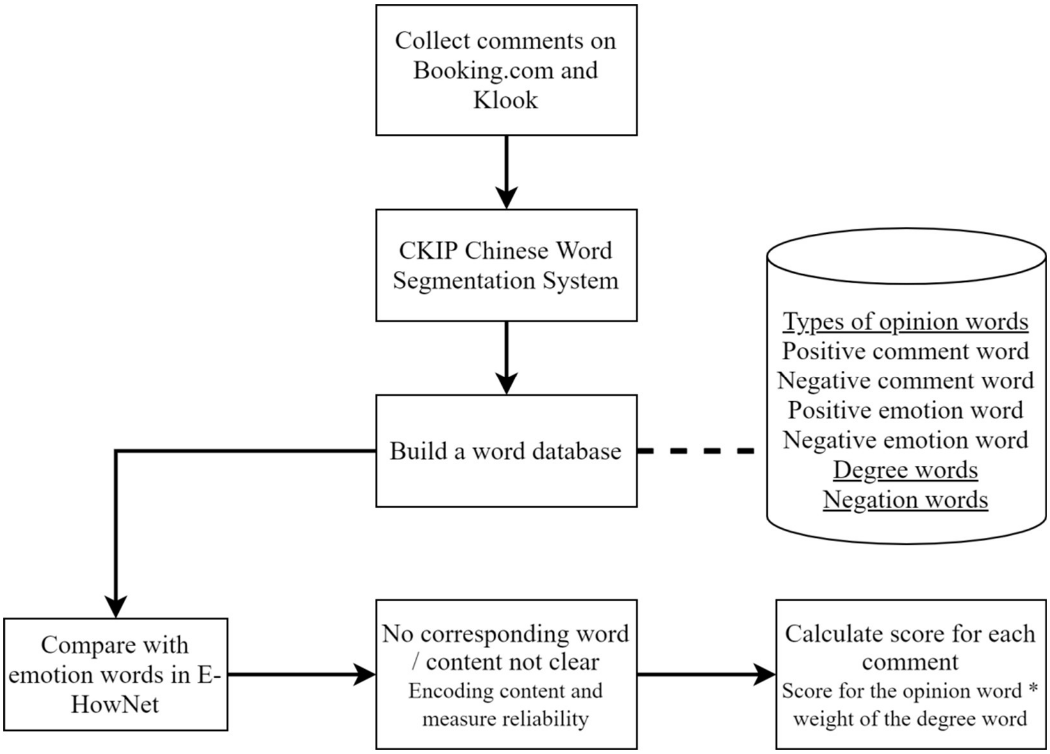

Figure 1

Method 1 analysis framework

(1) Online comments

In this study, we collected consumers' comments on Booking.com and Klook and classified them into three groups, including those on domestic hotels, foreign hotels, and one-day trips. According to Booking.com's statistics, the most popular domestic destinations in 2019 in Taiwan were Taipei, Kaohsiung, Taichung, and Tainan. Hence, in the collection of comments on domestic hotels, the scope of the data collection was set to cover only these four areas. As for comments on foreign hotels, Tokyo, Hong Kong, and Osaka have ever been the most popular overseas destinations among domestic Taiwanese tourists (https://travel.ettoday.net/article/184950.htm). Thus, comments were collected only from visitors to these cities. Comments on one-day trips were collected from Klook. Meaningless words were filtered out during comment collection.

(2) Chinese word segmentation system

The CKIP Word Segmentation System was used to segment the text of each comment and classify the segmented units by word class into verbs, nouns, prepositions, conjunctions, adverbs, auxiliaries, and adjectives. Take the comment “The counter service attendant is very professional” as an example: after word segmentation, it becomes “The counter (Nc) service (VC) attendant (Na) is very (Dfa) professional (VH)”.

(3) Building a dataset of words

Through the above-mentioned comment collection and word segmentation procedures, the words in the text were classified and a dataset of words could be obtained. Later, a sentiment dictionary, E-HowNet, was employed to create a series of thesauruses, including an opinion word thesaurus, a degree word thesaurus, and a self-compiled negation word thesaurus. In the collection of opinion words, words were classified by comment and emotion into four types, as shown in Table 1. A positive comment suggests that the consumer is sharing his/her positive feelings about the product or service, or satisfaction with its quality after consumption (Brown et al., 2005). A negative comment suggests that the consumer is sharing his/her negative feelings about the product or service, or dissatisfaction with its quality after consumption. Positive emotions can broaden our awareness and vision, facilitate the building of learning skills, and lead us to make contributions, while negative emotions may narrow our chances of survival (Zemach, 2001).

Table 1

Results of classification based on E-HowNet

Opinion type	Number of words	Examples
Positive comment	1893	Clean, comfortable, convenient, outstanding
Negative comment	1857	Dirty, boring, dark, stressful
Positive emotion	937	Ecstatic, joyful, elated, overjoyed
Negative emotion	825	Gloomy, angry, stunned, impatient

As for the collection of degree words, six degree words were adopted, including “very slightly”, “slightly”, “more”, “very”, “most”, and “extremely”.

The scores for these words were calculated using the formula of degree words weights and converted into 2.3, 3.8, 4.5, 4.8, 4.9, and 5, respectively. The higher the score, the higher the intensity. As for negation words, based on the collected comments, we utilized a total of 30 negation words or phrases to create a thesaurus of negation words, including nope, absent, unexpected, useless, failed, none, unable, not, negative, forbidden, unnecessary, poor, no, cannot, not enough, not allowed, no way, will not, impatient, indifferent, impossible, dislike, unclear, not seen, not notice, reject, unhappy, disrespect, inconvenient, and not specific. These negation words or phrases were intended to support detection of the polarity of words in the collected comments. This procedure was necessary because a negation word or phrase added before a positive comment may have resulted in the entire comment becoming negative.

(4) E-HowNet

We adopted E-HowNet as the thesaurus in this study. E-HowNet is a Chinese sentiment thesaurus based on the ontological structure and semantic definition mechanism of HowNet. It consists of 5512 words, each having a valence degree and an arousal degree. Valence refers to the degree to which an emotion word is positive or negative, while arousal indicates the degree of calmness or excitement. Both indices are measured on a scale from 1 (highly negative or calm) to 9 (highly positive or excited). This thesaurus was used to classify the collected set of words into opinion words and degree words for the subsequent sentiment analysis.

3.1.2 Content analysis and scoring of opinion words

In the sentiment analysis, the score for each sentiment word was determined based on the scoring method for E-HowNet. The comments extracted from Booking.com and Klook were segmented into words using the word segmentation system. Later, through a comparison with the emotion words in the sentiment dictionary, a score was derived for each word. However, in some situations there was no reference word for comparison. In these cases, the coding method was applied to create a semantic coding table. To ensure expert reliability, we invited three experts to review the coding results and measured the reliability through inter-rater agreement. Below is the formula used to test reliability:

M: Number of items fully agreed on by all raters

Ni: Number of items that Rater i should have agreed to;

Nj: Number of items that Rater j should have agreed to;

N: Number of raters;

Inter-rater reliability=2M /(Ni +Nj)

Reliability = n×mean inter-rater agreement / { 1+(n-1)×mean inter-rater reliability }

In scoring the opinion words, when a degree word was present, the score of the opinion word was multiplied by the weight of the degree word. The higher the score, the more the consumer cared about that aspect of the hotel or trip assessed in the comment; the lower the score, the less that the consumer cared. Hence, only opinion words and degree words were reserved for scoring.

3.2 Method 2: Decision tree analysis

3.2.1 Analysis framework

First, we administered a questionnaire to collect consumer data and responses for subsequent analysis and prediction. This questionnaire consisted of items across three major dimensions: demographics, the psychological dimension, and the behavioral dimension. The psychological dimension was used to evaluate consumers' lifestyles and attitudes with respect to the comments. To evaluate consumers' attitudes with respect to the comments, we used the comments obtained in Method 1 as the foundation and selected words with higher scores to design the items. Later, we carried out a reliability test, factor analysis, cluster analysis, and customer segmentation based on responses to items in the psychological dimension using SPSS Statistics. Further, we created a decision tree analysis using SPSS Modeler with three major market segmentation variables as predictor variables. Based on the decision tree analysis results, we attempted to predict consumer behavior in the target market to provide marketing suggestions. The analysis framework of Method 2 is as illustrated in Figure 2.

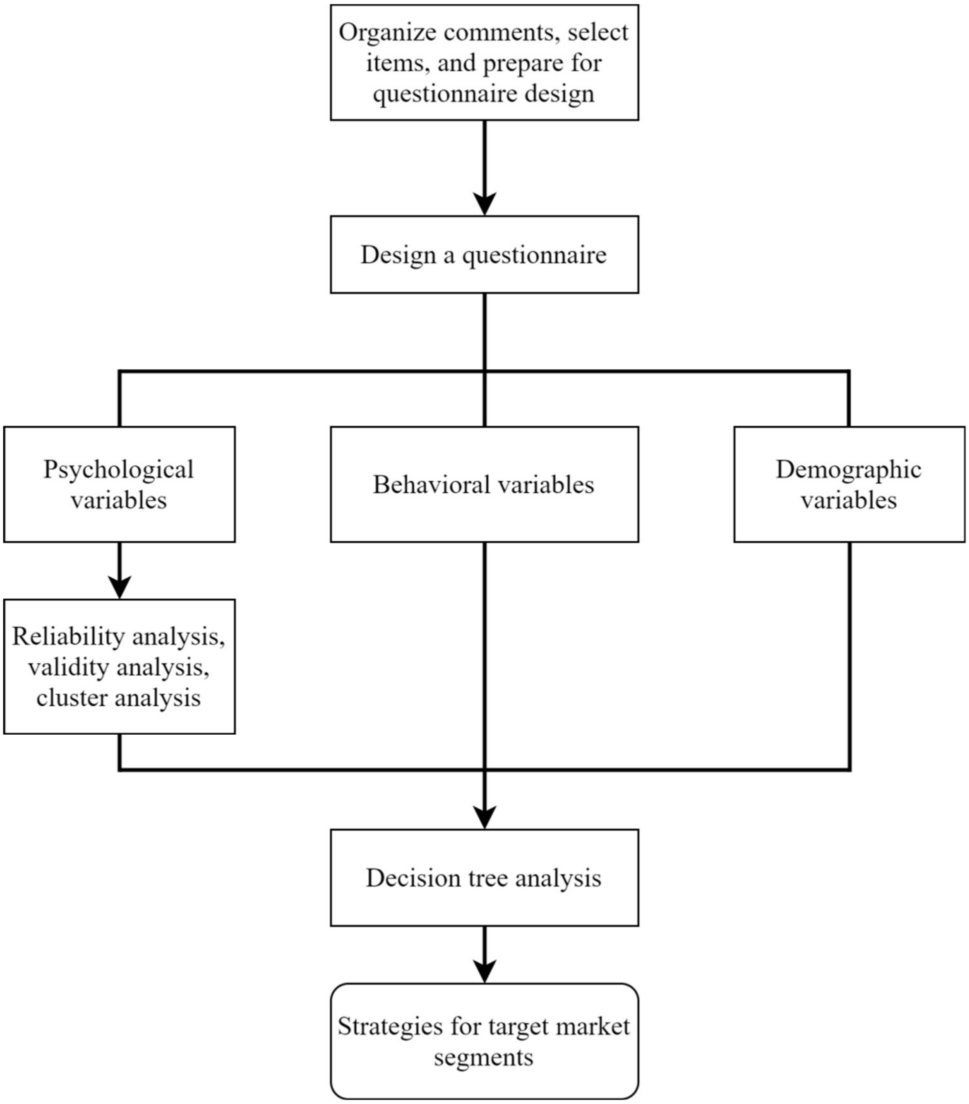

Figure 2
Method 2 analysis framework

The three dimensions of the questionnaire are explained as follows:

(1) Behavioral dimension

To survey consumers' travel-related experiences, we designed nine items based on those in the behavioral dimension. These items were used to collect data on purchase platforms, tourist destinations, tour types, spending each time, spending each year, travel goals, repurchase intention on the travel platforms, recommendation intention, and word-of-mouth intention.

(2) Psychological dimension

The psychological dimension encompasses lifestyles and customer comments. The Activities, Interests, and Opinions Scale (AIO) proposed by Peter and Olson (1994) is comprised of the three dimensions noted in the name. In addition, based on the lifestyle scales developed by Plummer (1974) and Reynolds and Williams (1973), we developed 21 items to measure lifestyles. These items were rated on 7-point Likert scales. As for customer comments, we used the words that appeared more frequently in the comments collected from the travel platforms as a reference to develop 22 items for measuring customer comments. These items were designed to be rated on a semantic differential scale.

(3) Demographic dimension

A total of seven personal profile variables were measured in this dimension: gender, age, highest level of education attained, occupation, marital status, family size, and average monthly income.

3.2.3 Factor analysis and cluster analysis

Through a factor analysis of data in the lifestyles and consumer attitudes in the comments, we deleted items with a factor loading below 0.4 and a cross loading (Hair, 2006). Later, we carried out a cluster analysis to cluster the participants by lifestyle and attitudes in the comment. The non-hierarchical K-mean clustering algorithm was adopted to reduce the required amount of data and computational effort and obtain clustering results that better reflect the reality of the situation.

4. Analysis results

4.1 Method 1: Sentiment analysis

4.1.1 Content coding and reliability analysis

In the coding process, words with a similar semantic meaning were given the same code. For those without a corresponding word in E-HowNet, we developed a semantic coding table and used inter-rater agreement as the reliability measure. We invited three experts in marketing management to evaluate the coding result. As shown in Table 2, we obtained a reliability result of 0.95 for the content coding and of 0.97 for the mean inter-coder agreement. According to Kassarjian (1977), reliability values greater than 0.85 are acceptable.

Table 2
Reliability analysis of the comment coding result

Inter-rater agreement	A	B
B	2*42/(44+44)≒0.95	
C	2*44/(44+44)=1	2*42/(44+44)≒0.95
Reliability	3*0.97*[1+(3-1)*0.97]≒0.95	

4.1.2 Comments

From Booking.com and Klook, we respectively collected comments about hotel accommodation and comments about travel itineraries that were made from November 1, 2019, to December 31, 2019. The collected comments included both positive and negative comments made by the target customers and key opinion leaders (KOL) on these two platforms. Later, we created a word cloud based on word frequency using WordClouds.com. The comments were analyzed as follows:

(1) Positive comments about accommodation

Comments about accommodation were obtained from Booking.com. After deletion of irrelevant and vague data, 294 comments remained. These comments were then categorized into room environment, facilities, geographic location, accessibility, attitude of the service attendants, and room price. Further, the total score for each comment, the mean score for each word, and the frequency of each word were also calculated. Among the positive comments about accommodation, “comfortable and cozy environment” had the highest mean score of 48.18. In terms of frequency, “clean environment” appeared 57 times, followed by “convenient transportation” (44 times), “nice and friendly service attitude” (36 times), “comfortable and cozy environment” (23 times), and “good value for money” (16 times). See the details in Table 3 and the word cloud in Figure 3. Moreover, we found that the KOLs’ comments (see the content in the red frame of Table 4) focused on the friendliness of service attendants, nice room, convenience, warmth, and new. The total scores for the major comment words are shown in Table 4.

Table 3
Average score for each positive word about accommodation

Comment category	Comment words	Mean score	Frequency (times)
Environment	Clean	35.04	57
	Comfortable and cozy	48.18	23
	Bright	31.68	6
	Spacious	23.07	6
	Quiet	26.12	10
Facility	All in readiness	31.68	6
Geographical location	Easily accessible	8.6	8
Transportation	Convenient	28.8	44
Service attitude	Nice and friendly	21.2	36
Price	Good value for money	7.8	16

Note. The higher the average score, the better the overall evaluation.

Table 4

The total score for each positive word in the comments (part of the data shown)

Opinion word / Score / Degree word	Clean (7.3)	Friendly (7)	Good (7.6)	Nice room (6.8)	Covenient (6)	Great (6.8)	Warm (7.4)	New (6.6)	Passionate (6.6)	Comfortable (6.6)	Recommend (6.2)	Comfort (6.6)	Total score
Very(4.8)					4.8*6								
Most(4.9)		4.9*7											84
	7.3						7.4				6.2		
			7.6										7.6
Very (4.8)	4.8*7.3												48.04
				6.8							6.2		
Very (4.8)						4.8*6.8							
Most (4.9)					4.9*6								62.04
Very (4.8)	4.8*7.3										4.8*6.2	4.8*6.6	96.48
Very (4.8)												4.8*6.6	61.08
Most(4.9)					4.9*6								
Very (4.8)						4.8*6.8	4.8*7.4						
Most(4.9)					4.9*6								97.56
Very (4.8)		4.8*7						4.8*6.6					
Most(4.9)				4.9*6.8									98.6
Very (4.8)	4.8*7.3									4.8*6.6			66.72
Very (4.8)	4.8*7.3									4.8*6.6			66.72
Very (4.8)				4.8*6.8		4.8*6.8		4.8*6.6			4.8*6.2		126.72
Very (4.8)						4.8*6.8				4.8*6.6			64.32
Very (4.8)						4.8*6.8				4.8*6.6			64.32
Very (4.8)	4.8*7.3				4.8*6								63.84
Very (4.8)	4.8*7.3				4.8*6			4.8*6.6					95.52
Very (4.8)						4.8*6.8			4.8*6.6				64.32
Very (4.8)				4.8*6.8	4.8*6								61.44

Figure 3
Word cloud of positive words about accommodation

Note. The positive words are arranged in order of frequency

(2) Negative comments about accommodation

Using the method noted above, we collected a total of 174 negative comments about accommodation. Among these comments, the word with the highest score was "dark" (18.24), followed by "noisy" (16.32); the word with the highest frequency was "uncomfortable" (50 times), followed by "small" (30 times), "noisy" (28 times), and "inadequate facilities" (15 times). See Table 5 for details and Figure 4 for the word cloud. In addition, the KOLs' comments (see the content in the red frame of Table 6) primarily focused on soundproofing and total room space, facility stability, hotel accessibility, and tidiness, as shown in Table 6.

Table 5
Average score for each negative word about accommodation

Comment category	Comment words	Mean score	Frequency (times)
Environment	Dirty	3.2	10
	Uncomfortable	9.8	50
	Dark	18.24	9
	Small	5.605	30
	Noisy	16.32	28
Facilities	Inadequate facilities (poor soundproofing)	9.5	15
Geographical location	Remote	8.5	5
Transportation	Not easily accessible	9.7	6
Service attitude	Poor attitude	11.5	5
Price	Unworthy	3.6	4

Note. The higher the average score, the worse the overall evaluation.

Table 6

The total score for each negative word in the comments (part of the data shown)

Opinion word / Score / Degree word	Poor soundproofing (3.4)	Poor attitude (3.6)	Bad smell (3.6)	Narrow (3.8)	Unstable (4.1)	Remote (4.5)	Noisy (3.6)	Careless (4.2)	Dirty (3.2)	Dark (3.7)	Inconvenient parking (4)	Not as advertised (3.8)	Total score
Slightly(3.8)	3.8*3.4												12.92
	3.4	3.6											7
Slightly (3.8)	3.8*3.4						3.6						16.52
	3.4				4.1					3.7			11.2
	3.4				4.1		3.6	4.2					15.3
Slightly (3.8)	3.8*3.4												12.92
Slightly (3.8)								3.8*4.2	3.8*3.2			3.8	31.92
Very(4.8)	4.8*3.4											3.8	20.12
More(4.5)	4.5*3.4					4.5*4.5							35.55
	3.4		3.6						3.2	3.7	4		10.6
	3.4				4.1								7.5
		3.6											3.6
		3.6		3.8									7.4
Slightly (3.8)			3.8*3.6		3.8*4.1								29.26
			3.6			4.5			3.2				11.3
Slightly (3.8)				3.8*3.8									14.44
Very(4.8)				4.8*3.8	4.1					4.8*3.7	4.8*4		59.3
				3.8		4.5							8.3
Slightly (3.8)				3.8*3.8									10.64
Very(4.8)				4.8*3.8	4.8*4.1								37.92

Figure 4

Word cloud of negative words about accommodation

Note. The negative words are arranged in order of frequency

(3) Positive comments about tours

We collected a total of 103 comments about travel itineraries from Klook and classified them into one of three categories: tour price, tour experience, and leader/guide. In this section, "city" became the variable to measure. In terms of tour price and tour experience, Osaka led all other cities with 45.36 and 73.36 points, respectively. In terms of leader/guide, Tokyo had the highest score with 27.3. See Table 7 for more details. As shown in Table 8, among the positive words in the comments, "recommend" appeared 16 times, "friendly" 20 times, and "professional" 16 times. The word cloud of the positive words is presented in Figure 5. In addition, the KOLs' comments (see the content in the red frame of Table 9) feature a greater emphasis on relaxing experience, friendly attitude and professionalism of the leader/guide. The total score table is shown in Table 9.

Table 7

Average score of positive comments for each destination

Comment category	Domestic	Mean score	Overseas	Mean score
Tour price	Taipei	22.3	Tokyo	38.84
	Taichung	25.8	Osaka	45.36
	Tainan	33.7	Hong Kong	25.48
	Kaohsiung	20.8		
Overall score		25.65		36.56
Tour experience	Taipei	34.2	Tokyo	63.75
	Taichung	38.3	Osaka	72.36
	Tainan	44.7	Hong Kong	34.5
	Kaohsiung	26.5		
Overall score		35.93		56.87
Leader/guide	Taipei	19.3	Tokyo	27.3
	Taichung	10.5	Osaka	26.8
	Tainan	25.2	Hong Kong	10.5
	Kaohsiung	20.3		
Overall score		18.83		21.53

Note. The higher the average score, the better the evaluation.

Table 8

Counts of each positive word about their tour

Comment category	Comment words	Counts
Tour price	Good value for money	9
Tour experience	Easy	6
	Fun	5
	Enriched	7
	As advertised	8
	Smooth	9
	Recommend	16
	Satisfied	5
Leader/guide	Friendly	20
	Professional	16
	Good attitude	10
	Fluent language skill	4

Table 9

The total score for each positive word about the tours (part of the data shown)

Opinion word / Score / Degree word	Easy (6)	Pleasant (7)	Great (6.8)	Worthy (6.4)	Not bad (6.3)	Recommend (6.2)	Nice experience (6.8)	Friendly (6.6)	Convenient (6)	Good value for money (8.4)	Detailed explanation (5.4)	Relaxing (6.2)	Total score
Very(4.8)	4.8*6											4.8*6.2	58.56
Extremely(5)						5*6.2				8.4			39.4
Most(4.9)				4.9*6.4									31.36
Very (4.8)			4.8*6.8										
													37.64
		7											
Very (4.8)					4.8*6.3					8.4			38.64
			4.8*6.8										
Very (4.8)													
													63.64
Extremely (5)						5*6.2							
Extremely (5)			5*6.8										
													34
Very (4.8)		4.8*7											
													33.6
Very (4.8)							4.8*6.8						32.64
Very (4.8)			4.8*6.8									4.8*6.2	
													77.4
								6.6		8.4			
Very (4.8)							4.8*6.8						
													38.84
						6.2							
Very (4.8)							4.8*6.8		4.8*6				61.44
	6				6.3								12.3
Extremely (5)	5*6								5*6				60
	6			6.4				6.6			5.4		18.4
Extremely (5)				5*6.4									
													38
									6				
Very (4.8)	4.8*6				4.8*6.3						4.8*5.4		84.96
Very (4.8)	4.8*6									4.8*8.4			69.12
Very (4.8)									4.8*6				23.04

Figure 5
Word cloud of positive words about tours

Note: The positive words are arranged in order of frequency

(4) Negative comments about tours

We obtained 103 negative comments about travel itineraries. As shown in Table 10, with respect to tour price, Tainan has the highest score of 25.8. In terms of tour experience and leader/guide, Hong Kong led other destinations with 22.2 and 25.7 points, respectively. As shown in Table 11, among the negative words, “rushed”, “unprofessional”, and “attitude” appeared relatively more frequently (13, 13, and 20, respectively). The word cloud is as illustrated in Figure 6. Besides, among the KOLs’ comments (see the content in the red frame of Table 12), value, actual quality, and smoothness were considered to be more important tour aspects.

Table 10

Average score of negative comments for each destination

Comment category	Domestic	Mean score	Overseas	Mean score
Tour price	Taipei	22.7	Tokyo	12.8
	Taichung	19.3	Osaka	15.3
	Tainan	25.8	Hong Kong	23.8
	Kaohsiung	25.7		
Overall score		23.375		17.3
Tour experience	Taipei	15.3	Tokyo	16.3
	Taichung	13.9	Osaka	14.4
	Tainan	12.3	Hong Kong	22.2
	Kaohsiung	18.3		
Overall score		14.95		17.63
Leader/ guide	Taipei	9.3	Tokyo	17.62
	Taichung	8.2	Osaka	18.8
	Tainan	8.7	Hong Kong	25.7
	Kaohsiung	10.32		
Overall score		9.13		20.71

Note. The higher the average score, the worse the evaluation.

Table 11

Counts of each negative word about their tour

Comment category	Comment words	Counts
Tour experience	Unworthy	7
	Rushed	13
	Boring	8
	Simple	6
	Dishonest	9
	Not smooth	5
	Not recommended	8
	Dissatisfied	5
Leader/ guide	Unreasonable	5
	Unprofessional	13
	Poor attitude	20
	Limited language skills	3

Table 12

The total score for each negative word about the tours (part of the data shown)

Degree word \ Opinion word Score	Rushed (3.3)	Boring (3.8)	Simple (4.3)	Dishonest (3.4)	Not smooth (3.6)	Unreasonable (2.6)	Poor attitude (3.4)	Dissatisfied (3.2)	Not recommended (3.6)	Short (4.6)	Poor itinerary (3.4)	Total score
Very (4.8)	4.8*3.3											15.84
Slightly (4.3)	4.3*3.3											
												19.09
			4.3									
Very (4.8)	4.8*3.3									4.8*4.6		37.92
Very (4.8)										4.8*4.6		
												25.28
								3.2				
Very (4.8)	4.8*3.3											
												19.64
		3.8										
				3.4			3.4		3.6			10.4
							3.4				3.4	6.8
Slightly (4.3)							4.3*3.4					14.62
Very (4.8)		4.8*3.8		4.8*3.4	4.8*3.6							51.84
Slightly (4.3)	4.3*3.3											14.19
			4.3	3.4		2.6			3.6			13.9
Very (4.8)	4.8*3.3											
												22.84
		3.8						3.2				
					3.6		3.4	3.2			3.4	13.6
	3.3					2.6		3.2				5.8
				3.4					3.6			
												25.254
Very (4.8)		4.8*3.8										
Very (4.8)					4.8*3.6						4.8*3.4	33.6
			4.3					3.2				7.5
	3.3										3.4	6.7
				3.4		2.6	3.4					9.4

Figure 6
Word cloud of negative words about tours

Note: The negative words are arranged in order of frequency

4.2 Method 2: Decision tree analysis

4.2.1 Sample structure

We designed two versions of the questionnaire, one for domestic travel and the other for overseas travel. The respondents were people with domestic or overseas travel experience who had previously made at least one comment related to their travels on any social website or travel e-commerce platform. The questionnaire was administered in two ways from Apr 30~May 20, 2020: using paper-and-pencil or online. We obtained a total of 316 responses, where 305 of them were valid (valid response rate of 96.8%). Of these responses, 154 concerned domestic travel and 151 overseas travel.

The valid sample for overseas travel consisted of 68 males and 86 females. In terms of travel type, 116 responses were related to self-guided tours, and 38 about group tours. Most of the respondents were between 25 and 44 years of age

(35.1%) and unmarried (64.3%). In addition, the majority of these respondents had a family size of 3-5 persons (70.1%). Almost a third of the respondents, about 31.8%, had an average monthly income below NT$20,000 (1US$≒30.5NT$). In terms of travel destination, 22.7% of the comments were about Tokyo, 27.9% about Osaka, 30% about Japan except Tokyo, and 19.4% about Hong Kong. The travel platforms they indicated experience using (multiple choice) included Booking.com (94 persons), Agoda (76 persons), Trivago (56 persons), Hotels.com (46 persons), HotelsCombined (19 persons), Expedia (15 persons), Airbnb (35 persons), Klook (38 persons), and KKday (36 persons).

The valid "domestic travel" sample consisted of 62 males and 89 females, aged mainly between 25 and 44 years old (45.7%), with most unmarried (61.6%). The majority had 3-5 persons in their family (72.2%). Just over a third of respondents, about 33.8%, had an average monthly income between NT$40,001 and NT$60,000. In terms of travel destination, 25.8% of the comments were about Taipei, 20.5% about Taichung, 22.5% about Tainan, and 17.3% about Kaohsiung. The travel platforms they noted experience using (multiple choice) included Booking.com (93 persons), Agoda (65 persons), Trivago (60 persons), Hotels.com (34 persons), HotelsCombined (17 persons), Expedia (8 persons), Airbnb (19 persons), Klook (40 persons), and KKday (30 persons).

4.2.2 Factor analysis and cluster analysis

We conducted a factor analysis and a cluster analysis of the questionnaire items about lifestyles and customer comments. Through the factor analysis, the factors in each dimension were reduced to three. Each factor has a Cronbach's α greater than 0.7, suggesting good item consistency and stability. In the cluster analysis, we obtained the best result from three clusters and named them. In the lifestyle dimension, Cluster 1 was named "Seeking travel information", Cluster 2 "Seeking travel experiences", and Cluster 3 "Seeking travel insights". In the customer comments dimension, Cluster 1 was named "Viewing staff professionalism as important", Cluster 2 "Viewing internal environment as important", and Cluster 3 "Viewing external environment as important". The cluster analysis results are presented in Tables 13 and 14. Finally, we incorporated the cluster analysis results into the decision tree analysis.

Table 13

Cluster analysis of lifestyles

Factor \ Cluster	Cluster 1	Cluster 2	Cluster 3
Factor 1: Seeking travel insights	0.5125	-0.50164	0.45578
Factor 2: Seeking travel information	0.71272	0.07371	-1.44602
Factor 3: Seeking travel experiences	-0.64722	0.55357	-0.36001

Table 14

Cluster analysis of customer comments

Factor \ Cluster	Cluster 1	Cluster 2	Cluster 3
Factor 1: Viewing internal environment as important	-0.39016	0.47392	0.57056
Factor 2: Viewing staff professionalism as important	0.44697	-0.25983	-0.90391
Factor 3: Viewing external environment as important	0.12614	-1.37446	0.89518

4.2.3 Variable selection and analysis result

We then utilized a decision tree analysis, with repurchase intention on the travel platform, recommendation intention, word-of-mouth intention, and average annual travel spending as the target variables, and the behavioral dimension, psychological dimension, and demographic dimension as the predictor variables. The selected variables in the behavioral dimension included travel type, average spending per trip, and travel goal; the selected psychological variables included seeking travel information, seeking travel experiences, seeking travel insights, viewing staff professionalism as important, viewing internal environment as important, viewing external environment as important; the demographic variables included gender, age, marital status, family size, and average monthly personal income. For “repurchase intention” as the target variable, we used CHAID cross-validation algorithm and created a decision tree with 10 fields. For “recommendation intention”, “word-of-mouth intention”, and "average yearly travel spending" as the target variable, C&RT algorithm was deemed to be optimal. We used 12, 13, and 13 fields to construct the decision trees, respectively. Refer to Table 15 for the details. The analysis results are explained below.

Table 15
The variables of travel experience

<table>
<tr><td colspan="4">Target variables</td></tr>
<tr><td colspan="4">Repurchase intention
Recommendation intention
Word-of-mouth intention
Average annual travel spending</td></tr>
<tr><td colspan="4">Predictor variables</td></tr>
<tr><td colspan="2">Behavioral dimension</td><td colspan="2">Options</td></tr>
<tr><td colspan="2">Travel type</td><td colspan="2">1. Self-guided; 2. Group tour</td></tr>
<tr><td colspan="2">Average spending per trip (most recent)</td><td colspan="2">1. Less than $5000; 2.$5000-10000; 3.$10001-24000; 4. $20001-30000; 5 More than $30000</td></tr>
<tr><td colspan="2">Travel goal</td><td colspan="2">1. Family travel; 2. Graduation trip; 3. Employee travel; 4. Travel with friends; 5. Solo travel</td></tr>
<tr><td colspan="4">Psychological dimension</td></tr>
<tr><td rowspan="3">Lifestyle</td><td>Seeking travel information</td><td rowspan="3">Attitude in the comment</td><td>Viewing staff professionalism as important</td></tr>
<tr><td>Seeking travel experiences</td><td>Viewing internal environment as important</td></tr>
<tr><td>Seeking travel insights</td><td>Viewing external environment as important</td></tr>
<tr><td colspan="2">Demographic dimension</td><td colspan="2">Options</td></tr>
<tr><td colspan="2">Gender</td><td colspan="2">1. Male; 2. Female</td></tr>
<tr><td colspan="2">Age</td><td colspan="2">1. 18 or younger; 2. 19-24 years old; 3. 24-44 years old; 4. 45-64 years old; 5. 65 or older.</td></tr>
<tr><td colspan="2">Marital status</td><td colspan="2">1. Single; 2. Married</td></tr>
<tr><td colspan="2">Family size</td><td colspan="2">1. 1-2 persons; 2. 3-5 persons; 3. 6 persons or more</td></tr>
<tr><td colspan="2">Average monthly personal income</td><td colspan="2">1. Less than $20,000; 2. $20,000-40,000; 3. $40,001-60,000; 4. $60,000-80,000; 5. More than $80,000</td></tr>
</table>

Note: All dollar figures in NT

(1) Target variable: repurchase intention

The predicted values of repurchase intention for the 12 nodes were 1.75, 4.464, 4.571, 4.897, 5.111, 5.25, 5.402, 5.517, 5.583, 6.071, 6.286, and 7 respectively (see Figures 4-5). We classified values equal to or higher than 5.583 into the high repurchase intention group, values equal to or lower than 4.897 into the low repurchase intention group, and values in between them into the medium repurchase intention group. The “high repurchase intention” group includes the predicted values of 5.583, 6.071, 6.286, and 7, which represent Terminal Nodes 10, 12, 3, and 11, respectively. The characteristics of consumers in each of these nodes were as follows: Terminal Node 10: Consumers aged 19-44, whose most recent tour was a self-guided tour, having three persons in their family, and viewing the external environment of accommodation as important. Terminal Node 12: Married consumers aged 45-64, whose most recent tour was a self-guided tour, with more than three persons in their family, and viewing external environment of accommodation as important. Terminal Node 3: female consumers whose most recent tour was a self-guided tour, with 1-2 persons in their family, and spending more than $10,000NT on each trip. Terminal Node 11: Unmarried consumer, aged 45-64, whose most recent tour was a self-guided tour, with more than 3 persons in their family, and viewing external environment of accommodation as important. The “medium repurchase intention” group includes the predicted values of 5.111, 5.25, 5.402, and 5.517, which represent Terminal Nodes 4, 2, 9, and 7, respectively. The consumer characteristics of each node were as follows: Terminal Node 4: Female consumers whose most recent tour was a self-guided tour, with 1-2 persons in their family, and spending no more than $10,000NT on each trip. Terminal Node 2: Consumers whose most recent tour was a group tour, and whose average monthly income was below $40,000NT. Terminal Node 9: Consumers whose most recent tour was a self-guided tour, with more than 3 persons in their family, viewing internal environment or staff professionalism as important, and with an average monthly income below $40,000NT. Terminal Node 7: Consumers whose most recent tour was a self-guided tour, with more than 3 persons in their family, viewing internal environment or staff professionalism as important, with an average monthly income above $40,000NT, and spending more than $30,000NT a year on travel. The "low repurchase intention" group included the predicted values of 1.75, 4.464, 4.571, and 4.897, which correspond to Terminal Nodes 6, 1, 5, and 8, respectively. The consumer characteristics of each terminal node were as follows. Terminal Node 6: Male consumers whose most recent tour was a self-guided tour, with 1-2 persons in their family, and who were good at expressing opinions about travel. Terminal Node 1: Consumers whose most recent tour was a group tour, with an average monthly income above $40,000NT. Terminal Node 5: Male consumers whose most recent tour was a self-guided tour, with 1-2 persons in their family, and with a tendency to seek travel information or tour experiences. Terminal Node 8: Consumers whose most recent tour was a self-guided tour, with more than 3 persons in their family, viewing internal environment or staff professionalism as important, with an average monthly income above $40,000NT, and spending no more than $30,000NT a year on travel. See Figure 7 for a full illustration.

(2) Target variable: recommendation intention

The predicted values of recommendation intention were 3.75, 4, 4.056, 4.182, 4.429, 4.696, 4.972, 5.2, 5.296, 5.548, 5.8, and 5.833 respectively (see Figures 4-6). We classified values equal to or higher than 5.296 into the high recommendation intention group, values equal to or lower than 4.182 into the low recommendation intention group, and values in between them into the medium recommendation intention group. The "high recommendation intention" group consisted of predicted values including 5.296, 5.548, 5.8, and 5.833, which represent Terminal Nodes 2, 12, 7, and 10, respectively. The consumer characteristics of each terminal node were as follows. Terminal Node 2: Consumers whose most recent tour was a domestic tour, who collect travel information, and whose travel goal was family travel or travel with friends. Terminal Node 12: Consumers whose most recent tour was an overseas self-guided tour, with more than 3 persons in their family. Terminal Node 7:

Consumers aged 19-44, whose most recent tour was an overseas group tour and who consider staff professionalism to be more important. Terminal Node 10: Female consumers whose most recent tour was an overseas self-guided tour, with 1-2 persons in their family. The “medium recommendation intention” group included predicted values of 4,429, 4.696, 4.972, and 5.2, which respectively represent Terminal Nodes 11, 3, 5, and 8. The consumer characteristics of each node were as follows. Terminal Node 11: Male consumers whose most recent tour was an overseas self-guided tour, with 1-2 persons in their family. Terminal Node 3: Consumers whose most recent tour was domestic, who enjoy tour experiences or expressing opinions about travel, who have an average monthly income above $40,000NT, and whose travel goal was solo travel or family travel. Terminal Node 5: Consumers whose most recent tour was a domestic tour, who enjoy tour experiences or expressing opinions about travel, and with an average monthly income below $40,000NT. Terminal Node 8: Consumers aged 45-64, whose most recent tour was an overseas group tour, who consider staff professionalism more important and tend to search for travel information. The “low recommendation intention” group was comprised of predicted values including 3.75, 4, 4.056, and 4.182, which correspond to Terminal Nodes 1, 9, 4, and 6, respectively. The consumer characteristics of each terminal node were as follows. Terminal Node 1: Consumers whose most recent tour was domestic, who seek travel information, and whose travel goal was solo travel. Terminal Node 9: Consumers aged 45-64, whose most recent tour was an overseas group tour, who pay more attention to staff professionalism, and who enjoy travel experiences or expressing opinions about travel. Terminal Node 4: Consumers whose most recent tour was domestic, who enjoy travel experiences or expressing opinions about travel, with an average monthly income above $40,000NT, and whose travel goal is travel with friends. Terminal Node 6: Consumers whose most recent tour was an overseas group tour, and who pay close attention to the internal and external accommodation environment. See Figure 8 for more details.

(3) Target variable: word-of-mouth intention

As shown in Figures 4-7, the predicted values of word-of-mouth intention are 3.667, 3.833, 4.182, 4.718, 5, 5.333, 5.533, 5.581, 5.6, and 6.1. We classified values equal to or greater than 5.581 into the “high word-of-mouth intention” group, values equal to or smaller than 4.718 into the “low word-of-mouth intention” group, and values in between the them into the “medium word-of-mouth intention” group. The “high word-of-mouth intention” group was comprised of values including 5.581, 5.6, and 6.1, which respectively represent Terminal Nodes 10, 2, and 7. The customer characteristics of each node were as follows. Terminal Node 10: Consumers whose most recent tour was overseas, paying more attention to external environment or staff professionalism, and with an average monthly income below $40,000NT. Terminal Node 2: Consumers whose most recent tour was domestic, who consider the external accommodation environment to be more important, and with 3 or more people in the family. Terminal Node 7: Female, tour experience-oriented consumers whose most recent tour was overseas, who view external environment or staff professionalism as more important, and with an average monthly income above $40,000NT. The “medium word-of-mouth intention” group consisted of predicted values including 5, 5.333, and 5.533, which represent Terminal Nodes 6, 5, and 8, respectively. The customer characteristics of each node were as follows. Terminal Node 6: Female customers whose most recent tour was overseas, who view external environment or staff professionalism as important, seek travel information or have travel insights, and with an average monthly income above $40,000NT. Terminal Node 5: Consumers whose most recent tour was overseas, paying more attention to internal environment, and with an average monthly income below $40,000NT. Terminal Node 8: Male consumers whose most recent tour was overseas, who view external environment or staff professionalism as more important, earn more than $40,000NT on average each month, and with 3-5 persons in their family. The “low word-of-mouth intention” group consisted of predicted values including 3,667, 3.833, 4.182, and 4.718, which represent Terminal Nodes 1, 9, 4, and 3, respectively. The consumer characteristics of each node were as follows. Terminal Node 1: Consumers whose most recent tour was domestic, who pay more attention to external environment, and with 1-2 persons in their family. Terminal Node 9: Male consumers whose most recent tour was overseas, who pay more attention to external environment or staff professionalism, earn more than $40,000NT on average each month, and with 1-2 or more than 6 persons in their family. Terminal Node 4: Consumers whose most recent tour was overseas, who view internal environment as more important, and with an average monthly income above $40,000NT. Terminal Node 3: Consumers whose most recent tour was domestic, and who view internal environment or staff professionals as more important. See Figure 9 for more details.

(4) Target variable: Average annual travel spending

As shown in Figure 10, the group with average annual travel spending below $10,000NT was characterized as spending less than $10,000NT on the most recent tour, aged 45-64, and with a travel information seeking orientation. In the group with an average annual travel spending from $10,001-20,000NT, those spending more than $10,000 on the most recent tour were characterized as having a monthly income below $40,000NT, being aged 19-44, with a travel goal of employee travel or travel with friends; those spending less than $10,000NT on the most recent tour were characterized by being aged 19-24. The group with an average annual travel spending of $20,001-30,000NT was characterized by spending less than $10,000NT on the most recent tour, being aged 45-64, an orientation towards seeking good tour experiences and expressing opinions, and paying more attention to external environment or staff professionalism. In the group with an average annual travel spending of $30,001-50,000NT, those spending more than $10,000NT on each tour were characterized by a monthly income below $40,000NT, being aged 19-44, with a travel goal of solo travel or family travel; those spending less than $10,000NT on each tour were characterized by a monthly income below $40,000NT and being aged 45-64. Finally, the group with an average annual travel spending above $50,000NT was characterized by an average spending on each tour exceeding $10,000NT and a monthly income above $40,000NT.

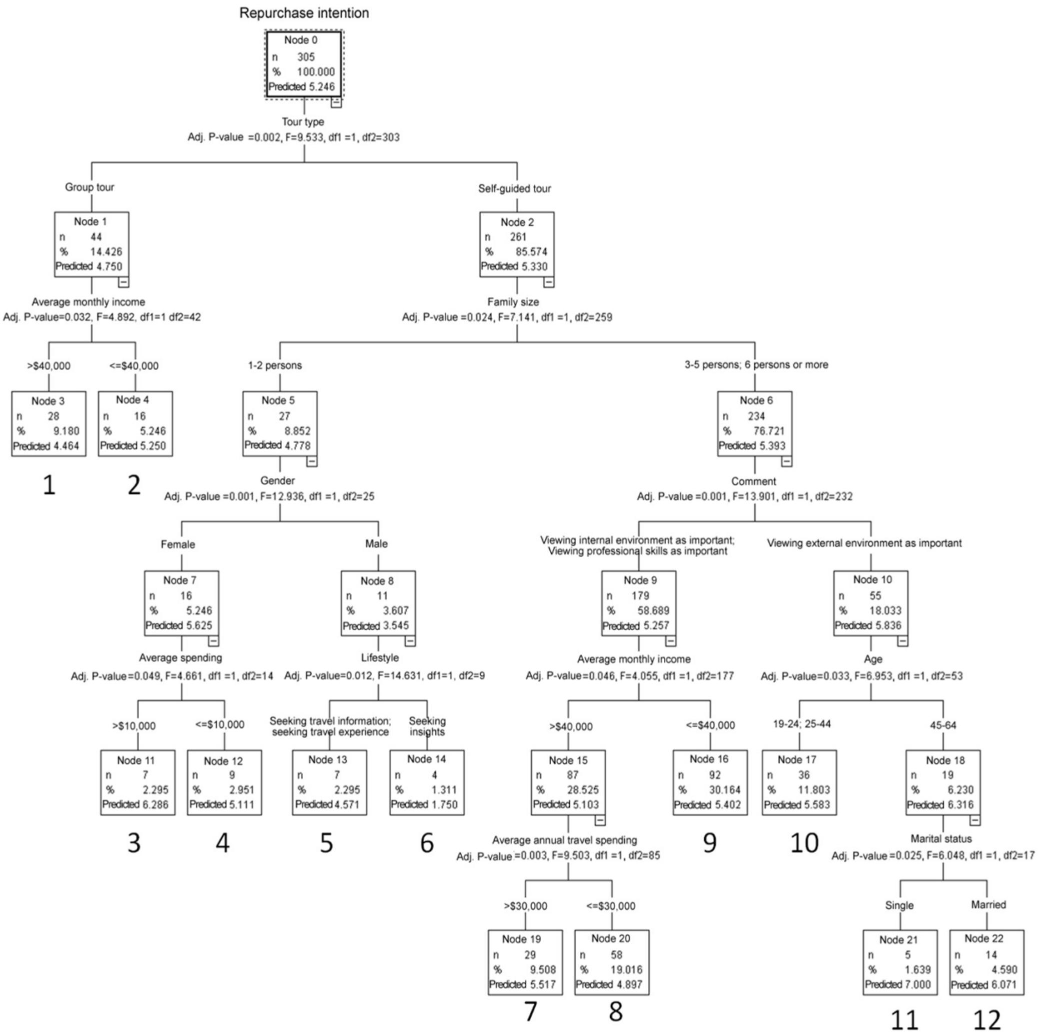

Figure 7

CHAID decision tree of repurchase intention

Note: Numbers 1, 2, 3…n indicate the terminal nodes

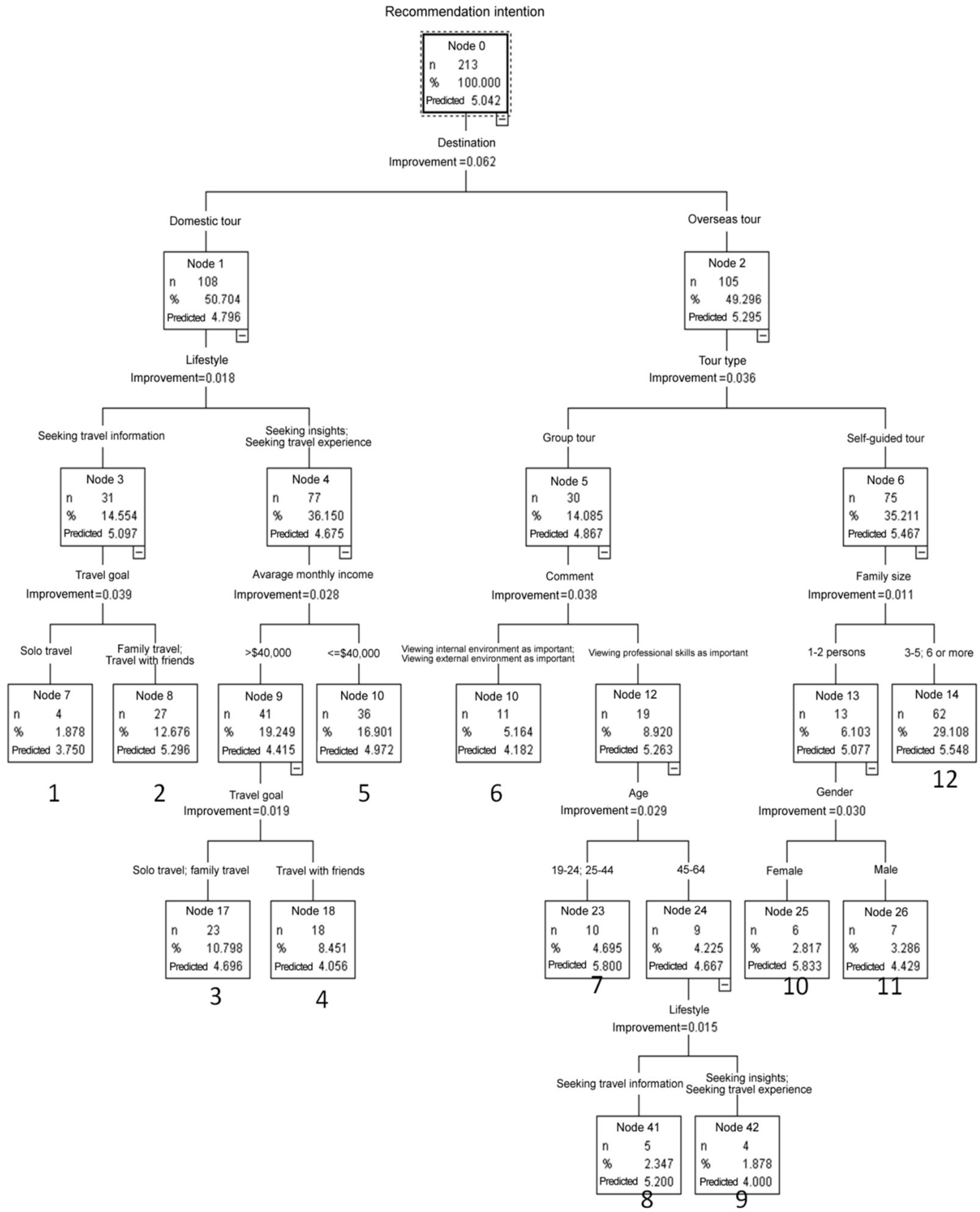

Figure 8

CandRT decision tree of recommendation intention

Note: Numbers 1, 2, 3…n indicate the terminal nodes

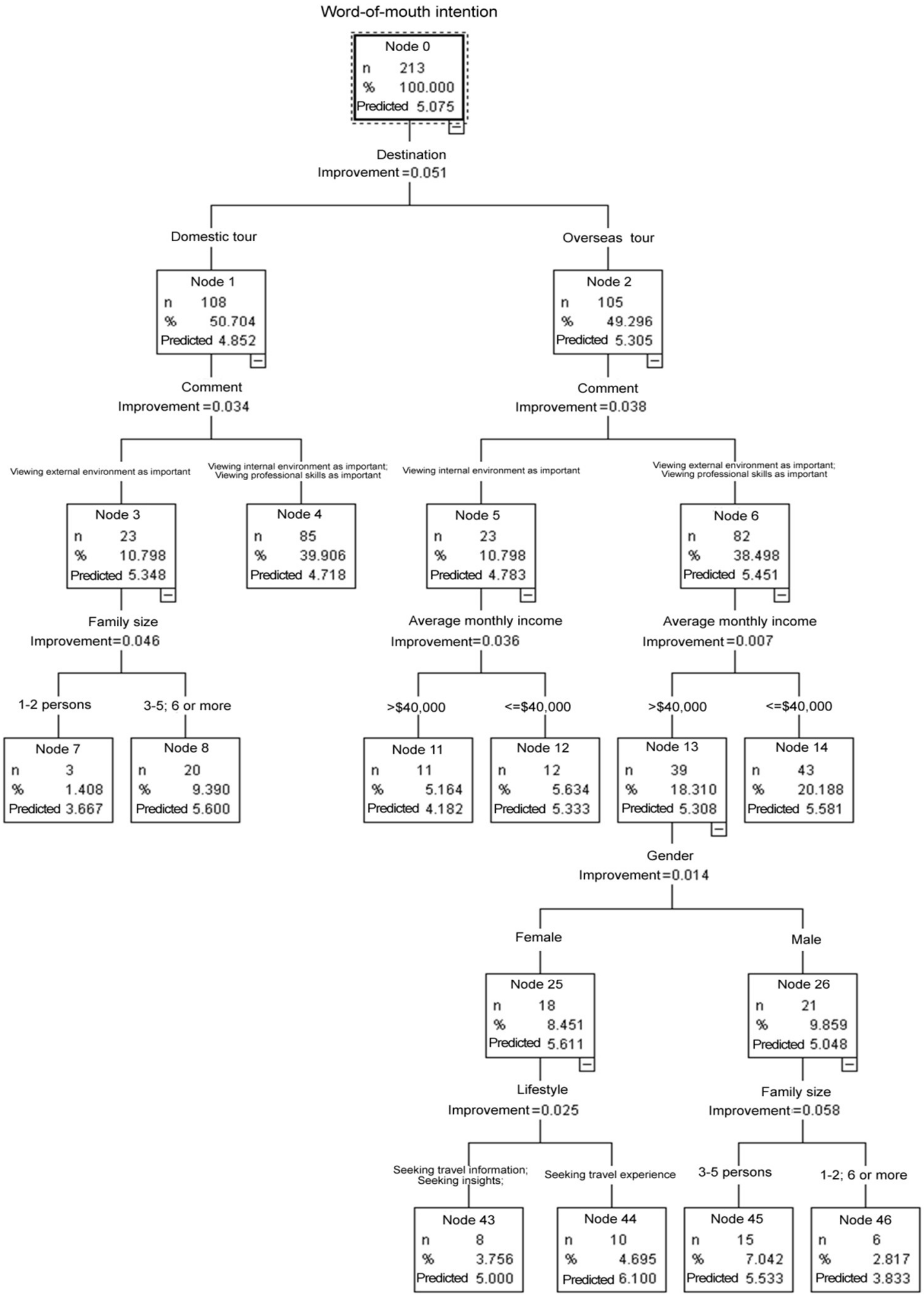

Figure 9

CandRT decision tree of word-of-mouth intention

Note: Numbers 1, 2, 3…n indicate the terminal nodes

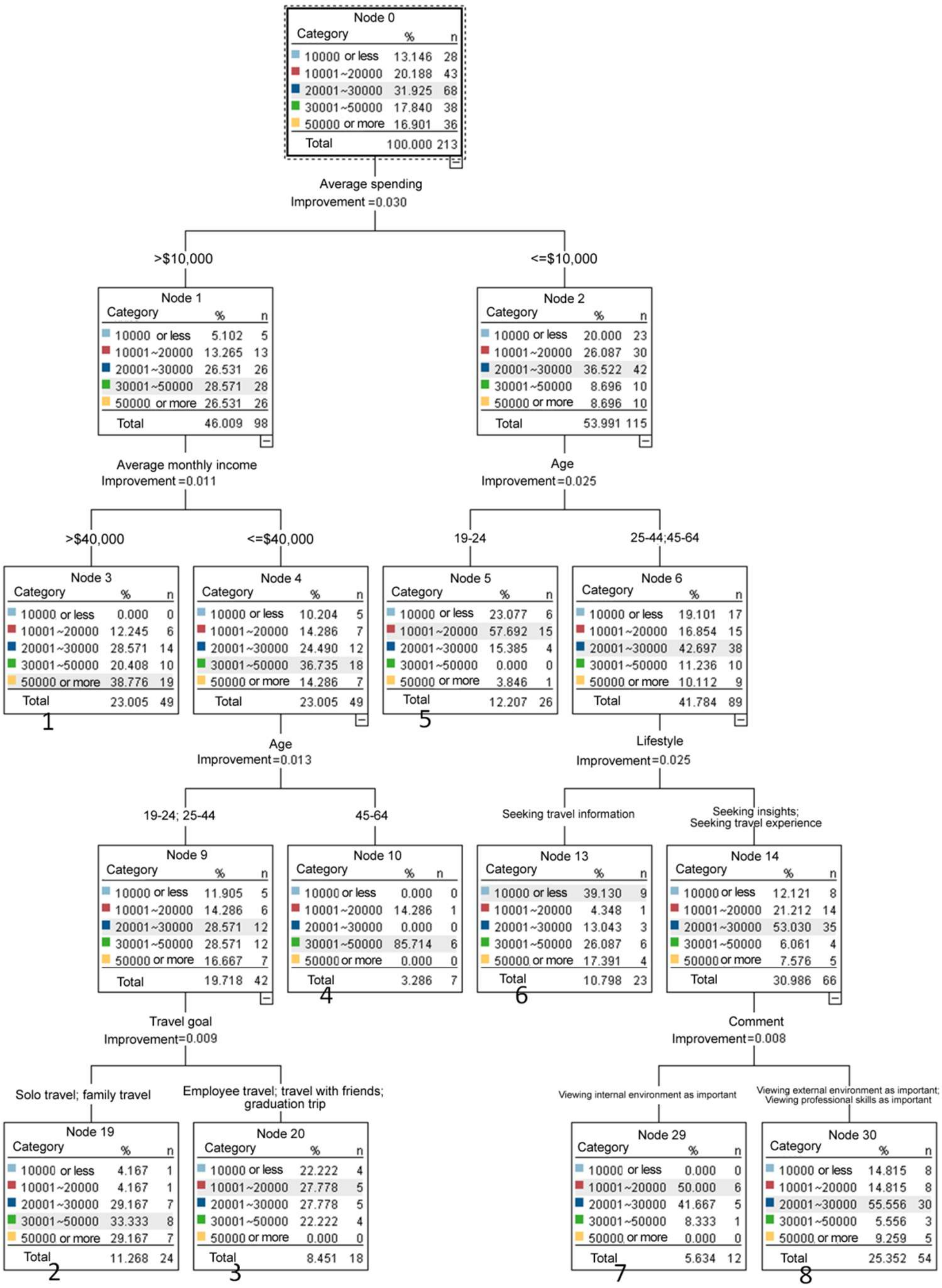

Figure 10

CandRT decision tree of average annual travel spending

Note: Numbers 1, 2, 3…n indicate the terminal nodes

5. Implications and conclusion

In this study, we apply sentiment analysis and decision tree analysis to cluster consumers with travel experiences and explore the effects of various variables. Based on the findings, we propose the following managerial implications, and then conclude with research limitations and future research suggestions.

5.1 Results

5.1.1 Sentiment analysis results

According to the analysis of comments collected from the travel platforms, in terms of the aspect of "accommodation", most comments, no matter positive or negative, concerned room cleanness, comfort and brightness, accessibility, space and facilities or service attitude. In these comments, consumers' sentiments about comfort, cleanness, and service attitude were particularly stronger, indicating that consumers care more about these details when staying in a hotel, and that these details affect their accommodation experience. In terms of the aspect of "tours", consumers tended to have more comments and opinions on the professional attitude of tour leaders and guides. In terms of positive comments, the score for overseas tours was higher no matter in the tour price, tour experience or leader/guide aspects. As for negative comments, domestic tours were associated with lower scores in terms of tour price, while overseas tours had lower scores in terms of both tour experience and the professional attitude of leader/guide. This suggests that consumers have different experiences and evaluations for overseas and domestic tours. When designing tours, service providers can factor in the above comments to create itineraries that better satisfy consumers' needs.

5.1.2 Decision tree analysis results

(1) Repurchase intention

In the decision tree analysis with "repurchase intention" as the target variable, travel type, family size, gender, average spending, lifestyle, customer comment, average monthly income, average annual travel spending, average travel spending per tour, age, and marital status were chosen as the segmentation variables. The results show that the predicted values of repurchase intention in most market segments exceeded the mean value, indicating that most consumers still have a high repurchase intention after making a purchase on travel platforms.

(2) Recommendation intention

In the analysis with "recommendation intention" as the target variable, segmentation variables including travel location, travel type, customer comments, age, lifestyle, average monthly income, travel goal, family size, and gender were considered. The results show that the predicted values of recommendation intention in most market segments were lower than the average, indicating most consumers have a lower intention to recommend products to family and friends after making a purchase on travel platforms.

(3) Word-of-mouth intention

The decision tree analysis with "word-of-mouth intention" as the target variable considered segmentation variables including travel location, comments, average monthly income, gender, lifestyle, and family size. The results indicate that the predicted values of word-of-mouth intention exceeding the average and those lower than the average made up almost an equal share of the market, suggesting that about half of consumers are willing to spread positive word-of-mouth for a travel platform or its service, while the other half has a lower intention to do so.

(4) Average annual travel spending

In the analysis with "average annual travel spending" as the target variable, we used average spending, average monthly income, age, travel goal, lifestyle, and customer comments as segmentation variables. We found that those having an annual travel spending of $20,001-30,000NT constituted the majority.

5.2 Managerial implications

Based on the results of sentiment analysis and decision tree analysis, we provide target marketing suggestions to travel platforms and accommodation providers.

5.2.1 Managerial implications of sentiment analysis

McCarthy (1960) proposed the 4Ps of marketing theory, suggesting that product, price, place, and promotion are the four pillars of a successful marketing project. Bitner and Booms (1981) extended the traditional 4Ps to 7Ps by including people, physical evidence, and process into the model. The 7Ps have become a basic framework of service marketing. In this study, we use the respondents' comments and perceptions of accommodation and tour products as the data and employ the 7Ps of marketing as the foundation to provide marketing suggestions to accommodation providers and travel platforms.

(1) Suggestions for accommodation providers

Product strategy: The suggested strategy covers physical products, services, and experiences. Accommodation room quality can be viewed as a kind of physical product. As shown in the word clouds, consumers tend to pay more attention to room tidiness, comfort, soundproofing, and overall hotel accessibility. Therefore, we suggest that accommodation providers enhance their cleaning of the internal and external spaces. For the aspect of comfort, they can provide more carefully selected or designed amenities, such as mouthwash, and dental floss. For the aspect of accessibility, they can provide more transportation information, taxi arrangements, or shuttle bus services. For the aspect of sound insulation, they can give customers a kind reminder, such as "Please lower your voices at night", or upgrade the soundproofing of their rooms.

Price strategy: The suggested price strategy covers the product's listed price, discounts, payment methods, and gifts or special offers. As shown in the word clouds, consumers view the accommodation cost-performance ratio and price as important. In turn, service providers can list different pricing for different room types, such as luxury suites, classical suites, and family rooms, allowing customers to choose the room that best fits their needs. Moreover, they can also set different prices for their rooms on different days, such as making a distinction between weekdays and weekend/national holidays or adopting a floating price for high and low seasons. They can also consider consumer tendencies, such as seeking lower prices and equating prices with quality, or consumer habits when setting the prices of their accommodation services. For example, they can apply mantissa pricing.

Place strategy: With the continuous growth of the travel industry and the expansion of its services, service providers can develop different marketing channels and partner with different types of booking platforms to give consumers more options to buy accommodation services. For example, they can use platforms such as Booking.com, Agoda, and Trivago to increase their market exposure and provide more comprehensive services information on the various platforms. Booking platforms can also utilize different marketing activities, such as Booking.com’s time limited offers and Agoda’s special offer on Monday, to attract the attention of different consumer groups.

Promotion strategy: Accommodation providers can leverage different means and tactics to achieve marketing goals and ensure that their target customers can better understand their products. They can offer special deals to attract consumers’ attention, such as accommodation and dining packages, a one night with two meals package, or tour packages that include visits to nearby tourists spots. Moreover, they can identify KOLs through an analysis of customers’ comments and attempt to schedule in-depth interviews with them. Once they understand the perceived weaknesses of their offerings, they should work to improve on them. Furthermore, they can invite KOLs to help promote their services, build positive impressions about and trust in their hotel, and attract potential customers.

People strategy: Service is the core of the travel industry. Explaining accommodation amenities to customers at check-in is a service. Controlling the impact of human factors can drastically reduce consumers’ complaints, increase the accommodation reputation and corporate image, and boost the service performance. As shown in the word clouds, consumers pay attention to the friendliness and enthusiasm of service attendants. Hoteliers can offer incentives to encourage employees to provide quality services. For example, they can hold a service excellence contest or offer training to all employees to enhance their work efficiency and quality, and also strengthen their identification with the company. When customers experience real staff enthusiasm, their favorability ratings are also likely to improve.

Physical evidence strategy: Physical evidence refers to the cues that consumers come into contact with, including the environment and service quality, which increase consumers' experiences with the hotel. As shown in the word clouds, consumers have strong sentiments about the accommodation novelty, warmth, and brightness. Service providers can use creative thinking and designs when constructing the interior of their hotel; when combined with excellent service quality, this can provide consumers with a better overall experience.

Process strategy: The service delivery process is one of the main elements of a service marketing mix. In the process of service delivery, enduring customer satisfaction cannot be achieved in only one step. In addition to the quality of the accommodation environment, the service delivery quality must be stressed. Enhancing these in terms of luggage storage, special services, and free services, among others, can contribute to greater use of the core product and increase the value-added to customers. Further, facilitating services, such as providing information, order taking, and payment handling, can also promote use of the core product. Service providers are suggested to pay extra attention to customer services before, during, and after service delivery and ensure consumer satisfaction throughout each stay. For example, they can provide detailed information before customers check in, offer as much assistance as possible to customers during each stay, and ensure smooth handling of customer payment during check-out. The weaknesses in each service process should be identified and carefully addressed.

(2) Suggestions for tour platforms

Product strategy: The products provided by platform operators are not physical. They are mainly the tour packages and activities across different countries and regions. However, as can be found in the word clouds, consumers pay close attention to the richness, comfort, and smoothness of a tour. Therefore, platform operators can develop more travel routes or new itineraries in different regions, and include activities that are unique to the region, such as culinary tours and old town tours. After an itinerary is designed, these operators should go through the entire trip themselves and make a detailed assessment of the length of stay at each attraction. The itinerary should be arranged in a way that avoids back-to-back activities and allows consumers to experience each attraction with ease and sufficient time.

Price strategy: For platform operators, price is not just a tag but also a signal of the service quality that customers should expect. Pricing strategies must be applied flexibly to ensure their effectiveness. As can be found in the word clouds, consumers are concerned about whether the price of a tour is good value for money. Therefore, platform operators can take advantage of their mindsets to seek business opportunities. Psychological pricing strategies, such as mantissa pricing, integer pricing, and prestige pricing can be applied to attract consumers. Moreover, price discounts, such as buy four and get one free or $300 off any purchase over $3000 can be used to stimulate consumption.

Place strategy: Due to intensifying competition in the tourism industry, travel platforms have no choice but to develop more diversified services. In addition to sale of itineraries, they can partner with companies in different industries. For instance, they can sell restaurant meal vouchers, offer special discounts on passes to amusement parks, sell accommodation vouchers or transport passes (e.g., flight tickets, ship tickets, and bus tickets) all through partnerships. In addition, they can also participate in various travel exhibitions to increase their popularity or cooperate with e-commerce platforms to launch time-limited offers and allow customers to buy their products in different ways.

Promotion strategy: To increase sales, tour platforms need to deliver the latest product information to consumers and apply marketing strategies to induce them to purchase. Operators are suggested to make use of combo sales such as buy Trip A and get Trip B for free, or include a gift with the product, such as buy a tour to Japan and get one power plug adapter for free. In advertising, they can utilize DM, online video ads, introductions by Internet celebrities, and TV ads to increase the media exposure of the platform. In addition, they should invite KOLs to join the initial group for a new itinerary, respond to KOLs' comments, and modify the itinerary based on their feedback. Because KOLs play an important mediator and filter role in mass communication, their recommendations can influence public discussion and draw the attention of potential customers.

People strategy: There are many important aspects of any tour offering in addition to the activities included in the itinerary, such as the attendants that customers encounter when shopping for a product, the tour leader or guide who runs the tour and the services they provide, and even the destination introduction and explanation. As shown in the word clouds, consumers place significant emphasis on the service attitude, professionalism, and language skills of tour leaders/guides. In terms of the service attitude aspect, platforms can analyze customers' opinions to identify the issues associated with each leader or guide, and then provide retraining or even stop working with those who resist retraining. For the professionalism aspect, they can use training to improve the professional skills of tour leaders and guides. With respect to the language skills aspect, this is an area of particular concern for overseas tour package consumers. When assigning a tour leader or guide to an overseas tour, platform operators should prioritize those proficient in multiple languages or having advanced skills in the language to be used during the tour. This can help them do a better job of explaining the attractions, as well as reduce misunderstandings caused by language problems, both of which will increase consumers' overall satisfaction with the tour.

Physical evidence strategy: Physical evidence can be divided into numerous aspects. For tour services, it includes the ambient conditions and the design of the tour; in the aspect of information communication, it includes service concretization and information concretization. As can be found in the word clouds, consumers often comment on the authenticity of trip content. Therefore, platform operators can make greater use of graphic and text content in designing web pages to quickly grab consumers' attention. They should provide more correct itinerary details to avoid confusion and reduce customer complaints caused by a gap between the actual itinerary and the information provided on the platform.

Process strategy: Service is variable in nature. The psychological states of service attendants and consumers may also vary depending on social or environmental factors. Therefore, we suggest that tour platforms regulate and normalize their services. For example, they can require their leaders/guides to: allow a minimum or fixed length of time of visit at each attraction, give an introduction of each attraction, and remain cognizant of where each participant is at all times; all of these are conducive to an improvement in service quality. Moreover, they can reassure consumers by providing detailed explanations of each itinerary on their website. For example, they can clearly indicate whether there are self-funded activities or shopping visits and provide more information about the leader and guide.

5.2.2 Managerial implications of decision tree analysis

(1) Maintaining and improving repurchase intentions

The policy for maintaining high repurchase intentions: The decision tree-based market segmentation analysis shows that customers with high repurchase intentions are characterized as female, having taken a self-guided trip as the most recent tour, spending more than $10,000NT on each trip, and paying more attention to the external environment (Terminal Node 3). Travel platforms can design a specific category of accommodation for this group of female customers. Operators can include accessibility, location, and surrounding attractions as search items when designing their platforms to appeal to those who view the external environment of accommodation as important. Accommodation providers can provide transport related services, such as shuttle bus rides or detailed direction information. For the itinerary aspect, service providers can design tour packages with attractions or activities that appeal to women, such as shopping, gourmet foods, and social-media hotspots. Targeting customers with high repurchase intentions is helpful for sales growth and market development.

The policy for improving low repurchase intentions: Customers with low repurchase intentions are characterized as male, having taken a self-guided trip as the last tour, with 1-2 persons in the family, and often expressing personal opinions about accommodations and tours (Terminal Node 6). For this group, platform operators can design customized short trips or recommend tour packages based on their needs. They can also provide more details on the accommodations or tour products to help this group find products that meet their needs. All these are conducive to development of a niche market.

(2) Maintaining and improving recommendation intentions

The policy for maintaining high recommendation intentions: The decision-tree based market segmentation analysis shows that customers with high recommendation intentions are characterized as having taken an overseas group tour as the last tour, paying attention to professional skills, and aged 19-44 (Terminal Node 7). Since this group of consumers is concerned about the professionalism and service attitude of the leader and guide travel platforms should prioritize those capable of speaking multiple languages fluently and with a strong service attitude. Customers will be more inclined to recommend a tour to their friends and relatives after they have had a satisfactory tour experience.

The policy for improving low recommendation intentions: The customers in this group are characterized as having taken a domestic tour as the last tour and with an interest in travel trends and information provided by travel websites (Terminal Node 1). For this group of consumers, platform operators can increase the information richness available through their websites, or organize activities on the platforms to attract attention and motivate them to share travel information with friends and relatives. Increasing the recommendation intention of this group of consumers is helpful for development of a niche market.

(3) Maintaining and improving word-of-mouth intentions

The policy for maintaining high word-of-mouth intentions: The decision tree-based market segmentation analysis shows that customers with high word-of-mouth intentions are characterized as having taken an overseas tour as the last tour, paying attention to the external environment and professional skills, earning more than $40,000NT a month, female, and having a regular travel habit (Terminal Node 7). This group of consumers view the direction information and transport services as very important and enjoy traveling while on holidays. For this group of consumers, platform operators can roll out more special offers on holidays, such as special hotel offers during the lunar new year holiday and special trips during consecutive holidays. Platform operators can also classify hotels and BandBs by service features to satisfy their need for regular travels.

The policy for improving low word-of-mouth intentions: The customers in this group are characterized as having taken a domestic tour as the last tour, paying attention to the external environment, and having 1-2 persons in the family (Terminal Node 1). Compared to overseas travel itineraries, domestic travel itineraries make up a much smaller proportion of those posted on travel platforms. We recommend that travel platforms offer a wide variety of domestic tours for small families, such as one-day Taipei trips or a lunar new year celebration in the national parks. Moreover, they can provide geographic and transportation information about the attractions to consumers. These are helpful for improving consumers' word-of-mouth intentions and creating a niche market.

(4) Implications from travel spending

Most of the consumers have an average annual travel budget of $20,001-$30,000NT. This group of consumers are characterized by spending less than $10,000NT on the most recent tour, being aged 45-64, having more personalized needs about accommodation and itineraries, having an interest in travel websites, and frequent travel. For this group of consumers, platform operators can offer mid-to-low priced tours and accommodations to attract their attention and send travel related information to them occasionally to boost their purchase intentions.

5.3 Theoretical implications

The results of this study contribute to the existing literature by providing empirical evidence on the role of consumer sentiment in shaping travel experiences and influencing decision-making in the tourism industry. Specifically, the sentiment analysis reveals that consumer satisfaction in the accommodation and tour segments is highly dependent on service quality factors, such as room comfort, cleanness, and service attitude. These findings support the theoretical frameworks of service quality and consumer satisfaction, indicating that emotional responses to specific service attributes significantly affect overall travel experiences. Additionally, the differentiation between consumer sentiments toward domestic and overseas tours underscores the importance of context-specific factors in shaping consumer perceptions, providing a theoretical basis for understanding cross-border travel behavior. This supports the theory that consumer evaluations are influenced by the service environment and cultural expectations, emphasizing the need for travel providers to adapt services based on geographic and cultural contexts.

This study extends academic discourse by incorporating decision tree analysis to examine behavioral intentions, such as repurchase, recommendation, and word-of-mouth intentions. The findings demonstrate that lifestyle, spending patterns, and demographics play crucial roles in predicting these intentions, offering a novel contribution to segmentation theory in tourism research. Furthermore, the contrasting results for recommendation intention (generally low) and repurchase intention (generally high) highlight a unique consumer behavior pattern: while customers are likely to make repeat purchases, they may be less inclined to recommend travel services to others. This discrepancy invites future research to explore the psychological factors behind recommendation reluctance despite overall satisfaction. The integration of big data-driven sentiment analysis and decision tree models provides methodological innovation, advancing the application of machine learning techniques in tourism research and offering a replicable framework for similar studies in other service-oriented industries.

5.4 Limitations and suggestions

Our methodology combines qualitative sentiment analysis and quantitative decision tree analysis, which is seldom discussed or applied in the tourism marketing and management area. This study is subject to a number of limitations, specifically regarding comment collection, sampling, analytic tools, and data sources. This section summarizes the limitations and provides future research suggestions.

(1) The comment collection scope

For the qualitative sentiment analysis, we collected comments about only a few popular tourist cities that were published during a period of two months, possibly affecting the results. Future researchers can include more cities and use a longer sampling period to collect more diversified comments for analysis.

(2) The consumer sampling scope

For the quantitative decision tree analysis, due to time and cost considerations, we were unable to collect a larger sample consisting of diversified demographics and different travel types. For example, the sample of group tour consumers was small, so the number of consumers in each layer of the decision tree was also small, making it harder for a variable to become a segmentation variable. Future researchers can work with travel platforms and their members to perform a more accurate market survey and market segmentation.

(3) Analytic tools

In the qualitative sentiment analysis, we adopted E-HowNet to compare the words. The amount of reference words for some words was insufficient. For these words, we carried out a content analysis and ensured the expert reliability and validity of the coding results. In the quantitative decision tree analysis, we adopted the CandRT and CHAID algorithms. Future researchers should apply different algorithms for different sample types and sizes to obtain increased accuracy and an optimal number of classification fields.

(4) Diversified data sources

We used big data in the travel industry as the research subject and analyzed comments about accommodations on Booking.com and comments about travel itineraries on Klook. Future researchers can include other social websites, such as TripAdvisor, Facebook, or travel forums to collect comments from multiple channels or to recruit participants for a questionnaire survey.

References

Adedoyin-Olowe, M., Gaber, M. M., and Stahl, F. (2014). *A survey of data mining techniques for social media analysis. Journal of Data Mining and Digital Humanities, 18.* https://doi.org/10.46298/jdmdh.5

Alaei, A. R., Becken, S., and Stantic, B. (2019). Sentiment analysis in tourism: *Capitalizing on big data. Journal of Travel Research,* 58(2), 175–191. http://10.1177/0047287517747753

Baars, H., and Kemper, H. G. (2008). Management support with structured and unstructured data-an integrated business intelligence framework. *Information Systems Management, 25(2),* 132-148. https://doi.org/10.1080/10580530801941058.

Banu, G. R. (2016).Predicting thyroid disease using linear discriminant analysis (LDA) *data mining technique. Communications on Applied Electronics, 4*(12), 4-6.

Basiri, M. E., Nemati, S., Abdar, M., Asadi, S., and Acharrya, U. R. (2021). A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. *Knowledge-Based Systems,* 228, Article 107242.

Bitner M. J., and Booms, B. H (1981). *Marketing strategies and organization structure for service firms.* In J. H. Donnelly and W. R. George (Eds), Marketing of Services (pp. 47- 52). Chicago, IL,: American Marketing Association.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. I. (1984). *Classification and regression trees.* Belmont, Calif.: Wadsworth.

Brooker, P., Barnett, J., and Cribbin, T. (2016). Doing social media analytics. *Big Data and Society 3*(2), 1-12. https://doi.org/10.1177/2053951716658060.

Brown, T. J., Barry, T. E., Dacin, P. A., and 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,* 123-138. https://doi.org/10.1177/0092070304268417.

Chang, M., Cheung, W., and Tang, M. (2013). Building trust online: Interactions among trust building mechanisms. *Information and Management,* 50(7), 439-445. https://doi.org/10.1016/j.im.2013.06.003

Chauhan, D. and Thakur, J. (2014). Boosting decision tree algorithm for weather prediction. *Journal of Advanced Database Management and Systems, 1*(3), 21-30.

García-Palomares, J. C., Gutiérrez, J., and Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography,* 63, 408-417. https://doi.org/10.1016/j.apgeog.2015.08.002.

Chen, H. A., Marmorstein, H., Tsiros, M., and Rao, A. R. (2012). When more is less: The impact of base value neglect on consumer preferences for bonus packs over price discounts. *Journal of Marketing, 76*(4), 64-77. https://doi.org/10.1509/jm.10.0443

Chen, H., Chiang, R. H. L., and Storey, V. C. (2012).Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems, 36*(4), 1165-1188. https://doi.org/10.2307/41703503

De Fretes, D. R., Noviany, H., Juliarta, E. F., Sandya, D., and Charviandi, A. (2024). Sentimental analysis of big data to measure tourism's social and emotional impact on social media: A case study of Indonesia. *International Journal of Society Reviews, 2*(1), 46–58. https://doi.org/10.56133/ijoser.v2i1.139

De Fretes, L. S., Hartono, P., and Suryanto, A. (2023). Sentiment analysis of social media data for assessing tourism's emotional impact in indonesia. *International Journal of Social and Economic Research, 11*(1), 45–62. http://10.1234/injoser.joln12345

Ellison, N. B., Steinfeld, C., and Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Fan, S., Lau, R. Y. K. and Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing. mix. *Big Data Research, 2*(1), 28-32. https://doi.org/10.1016/j.bdr.2015.02.006

Fang, X., Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data, 2*(1), 1-14. https://doi.org/10.1186/s40537-015-0015-2

Fronzetti Colladon, A., Giudici, E., and Grippa, F. (2021). Forecasting tourism demand with sentiment analysis: Insights from online travel communities. *Tourism Management Perspectives, 38,* 100819. http://10.1016/j.tmp.2021.100819

Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

García-Palomares, J. C., Gutiérrez, J., and Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography, 63*, 408-417. https://doi.org/10.1016/j.apgeog.2015.08.002.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., and Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21*(4), 431-454.

Hartigan, J. A., 1975. *Algorithm CHAID*. Clustering Algorithms. John Wiley and Sons, New York.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S. Gani, A., Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115. https://doi.org/10.1016/j.is.2014.07.006.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P. and Hawelka,

C. R. (2014). Geo-located Twitter as proxy for global mobility patterns. Geo-located. *Cartography and Geographic Information Science, 41*(3), 260-271. https://doi.org/10.1080/15230406.2014.890072.

IDCiView (2015). *Worldwide semiannual big data and analytics spending guide.* Needham, MA: IDC

Jena, R. (2020). An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Industrial Marketing Management, 90*, 605-616.

Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Kassarjian, H. (1977). Content analysis in consumer research. *Journal of Consumer Research, 4*(1), 8-18. https://doi.org/10.1086/208674.

Khan, M. A., Uddin, M. F. and Gupta, N. (2014). *Seven V's of big data understanding big data to extract value*. In IEEE (Ed), Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education (pp. 1-5). CT: Bridgeport https://doi.org/10.1109/ASEEZone1.2014.6820689.

Kim, J. and Hastak M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management, 38*, 86-96.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety.* META Group Inc. Web site: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Lee, C. K., Lee, Y. K., and Wicks, B. E. (2004). Segmentation of festival motivation by nationality and satisfaction. *Tourism Management, 25*(1), 61-70. https://doi.org/10.1016/S0261-5177(03)00060-8

Lu, R., Zhu, H., Liu, X., Liu, J. K., and Shao, J. (2014). Toward efficient and privacy-preserving computing in big data era. *IEEE Network, 28*(4), 46-50, https://doi.org/10.1109/MNET.2014.6863131.

Lv, Z., Song, H., Basanta-Val, P., Steed, A., and Jo, M. (2017). Next-generation big data analytics: State of the art, challenges, and future research topics. *IEEE Transactions on Industrial Informatics, 13*(4), 1891-1899. https://doi.org/10.1109/TII.2017.2650204.

Lynd D, Bacon (2002). *Marketing, handbook of data mining and knowledge discovery*. Edited by Willi Klosgen and Jan M. Zytkow, OXFORD University Press. 715-725.

Manovich, L. (2011). *Trending: The promises and the challenges of big social data.* In M. K. Gold (Ed.), Debates in the Digital Humanities. MN: The University of Minnesota Press. https://doi.org/10.5749/minnesota/9780816677948.003.0047.

Marine-Roig, E. and Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing and Management, 4*(3), 162-172. https://doi.org/10.1016/j.jdmm.2015.06.004.

Marshall, A., Mueck, S. and Shockley, R. (2015). How leading organizations use big data and analytics to innovate. *Strategy and Leadership, 43*(5), 32-39. https://doi.org/10.1108/SL-06-2015-0054.

McCarthy, E. J. (1960). *Basic marketing: A managerial approach.* Homewood, IL: Irwin.

McCleary, K. W. (1995). Applying internal marketing techniques for better festival organization and management. *Festival Management and Event Tourism, 3*(1), 1-7.

Meenaa, G., Mohbey, K. K., and Kumar, S. (2023). Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach. *International Journal of Information Management Data Insights, 3*(1), Article 100174.

Morabito, V. (2015). *Big data and analytics: Strategic and organizational impacts*. International Publishing. https://doi.org/10.1007/978-3-319-10665-6.

Ohbe, T., Ozono, T. and Shintani, T. (2019). A sentiment polarity classifier for regional event reputation analysis. *International Journal of Service and Knowledge Management International Institute of Applied Informatics, 3*(1), 51-65.

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., and Belfkih, S. (2018) Big data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences, 30*(4), 431-448. https://doi.org/10.1016/j.jksuci.2017.06.001.

Park, G., Chung, L., Khan, L., and Park, S. (2017). A modeling framework for business process reengineering using big data analytics and a goal-orientation. *Proceedings of the 2017 11th International Conference on Research Challenges in Information Science (RCIS),* 21-32. https://doi.org/10.1109/RCIS.2017.7956514.

Peter, J. P. and Olson, J. C. (1994). *Understanding consumer behavior*. Burr Ridge, IL: Irwin.

Plummer, J. T. (1974). The concept and application of life style segmentation. *Journal of Marketing, 38*(1), 33-37. https://doi.org/10.2307/1250164

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81-106. https://doi.org/10.1007/BF00116251.

Quinlan, J. R. (1993). C4.5: *Programs for machine learning.* Morgan Kaufmann Publishers.

Rao, V. R., and Kartono, B. (2009). Pricing objectives and strategies: A cross-country survey. In V. R. Rao (Ed.), *Handbook of pricing research in marketing* (pp. 9–36). Edward Elgar Publishing. https://doi.org/10.4337/9781848447448.00009

Reynolds, F. D., and Darden, W. R. (1973). Constructing life style and psychographics. In W. D. Wells (Ed.), *Life style and psychographics* (pp. 71–95). American Marketing Association.

Rumelhart, D. E., Widrow, B., and Lehr, M. A. (1994). The basic ideas in neural networks. *Communications of the ACM, 37*(3), 87+. https://link.gale.com/apps/doc/A15061351/AONE?u=anon~ea0d6afdandsid=googleScholarandxid=a7645bbf

Scott's Directories (2002). *Scott's business directory compile a list of companies.* Web site: https://www.scottsdirectories.com/.

Shaw, G and Williams, A. (2009). Knowledge transfer and management in tourism organisations: An emerging research agenda. *Tourism Management, 30*(3), 325-335.

Siegel, C. F. (2004). *Internet marketing: Foundations and applications.* Houghton Mifflin Company.

Sivarajah. U., Kamal, M. M., Irani, Z., and Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263-286. https://doi.org/10.1016/j.jbusres.2016.08.001

Sorensen, F., Jensen, J. F. (2015). Value creation and knowledge development in tourism experience encounters. *Tourism Management, 46*, 336-346.

Srivastava, A., Han, E. H., Kumar, V., and Singh, V. (1999). Parallel formulations of decision-tree classification algorithms. *Data Mining and Knowledge Discovery, 3*(3), 237-261. https://doi.org/10.1023/A:1009832825273.

Sun, H., and Hu, X. (2017). Attribute selection for decision tree learning with class constraint. *Chemometrics and Intelligent Laboratory Systems, 163*(15), 16-23.

Thomas, A. and Frank, H. J. (2015). Measuring the Degree of Corporate Social Media Use. *International Journal of Market Research, 57*(2), 257-275.

Tsai, K. (2016). *Five charts to understand Taiwan's travel websites, Travel Research Institute*. Web site: https://travel20.blogspot.com/2016/05/blog-post_27.html

Wamba, S. F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015). How "big data"can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 65*, 234-246. https://doi.org/10.1016/j.ijpe.2014.12.031.

Wu, X., Zhu, X., Wu, G. Q., and Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1) 97-107, https://doi.org/10.1109/TKDE.2013.109.

Zemach, E. (2001). What Is Emotion? *American Philosophical Quarterly, 38*(2), 197-207. http://www.jstor.org/stable/20010033.