Journal of Architectural/Planning Research and Studies
(JARS) Vol.21(1).2024

https://doi.org/10.56261/jars.v21.256845

Boutique Hotel Service Digitalization: A Business Owner Study

Somatat Na Takuatung[1]*, Chokeanand Bussracumpakorn[2]
Faculty of Architecture and Design, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
* Corresponding author e-mail: somatat.na@mail.kmutt.ac.th
Received 9/06/2022 Revised 12/10/2022 Accepted 19/10/2022

Abstract

The COVID-19 pandemic has generated negative, economic impacts on the tourism and leisure sector in Thailand, especially small boutique hotels. These hotels have had to develop more efficient and innovative approaches to meet new normal expectations, for example, contactless service. Digital technologies, such as Machine Learning and Artificial Intelligence, can open new possibilities and opportunities for hotels to digitize their customers' services. A review of the literature indicated that data important to the management of hotel products and services include Customer Segmentation, Customer Profiling, Menu Engineering, Productivity Indexing, Customer Associations, Forecasting, Energy Consumption, and Room Rates. These characteristics can be examined by machine learning. This study used a mixed qualitative and quantitative research method. The data were gathered by interviewing two boutique hotel owners in Bangkok and collecting the hotels' data, including online travel booking agents and direct booking logs, for the period April 2016 – September 2021. The analysis was conducted using the booking data from the two hotels: 3946 records from Hotel A and 3948 from Hotel B. In this research, k-means clustering was used to segment hotel guests. Two-class logistic regression and a two-class boosted decision tree were used to predict the prospective customer, while linear regression and decision forest regression were used to forecast the market demand. The findings reveal a model of hotel business owners' requirements to innovate new service solutions, such as the contactless software solution, that guests can employ for check-in, check-out, order services, and talk to the hotel through the mobile application. This would help hotel owners to manage costs, employees, and customers. The solution also means that hotel managers would no longer need to be involved in the manual implementation of revenue management tasks. This data analytics approach can effectively sift through the signals detected from market variables, discover patterns and anomalies, make predictions for guest arrivals, and calculate optimum prices in real-time, as the market changes.

Keywords

Hotel Service Digitalization; Service Design; Mobile Application; Hotel Industry; Contactless Service

1. Introduction

1.1 Background

According to Industry Outlook 2021-2023 Hotel Industry (Lunkam, 2021), the hotel business (which here includes hotels, resorts, and guesthouses) is directly connected to and an important part of the wider tourism sector. In terms of its contribution to the economy, in 2019, accommodation and food service activities together accounted for 6.1 percent of Thai gross domestic product (GDP), bringing in 1.03 trillion Thai Baht.

The sector has a central role in the overall economy because Thailand is one of the world's popular tourist destinations. This is partly because the nation the world-class tourist attractions are found throughout the country. Bangkok is a major and perennially popular tourist attraction, as underlined by the tourism awards that the city consistently wins. There are also famous seaside and beach destinations in the South and East, and a range of ecotourism travel options in the North. In addition, the country benefits from competitive pricing for accommodation and a low cost of living, which offers better value-for-money than many other countries. Beyond this, the travel industry benefits from the country's extensive, comprehensive transportation network, national infrastructure that is constantly being upgraded, and the increasing number of low-cost carriers serving the local market. These factors helped to give Thailand an edge over its competitors. Indeed, the 2019 Travel and Tourism Competitiveness Index, compiled by the World Economic Forum places Thailand in 31st place out of 140 countries surveyed (up from 34th in 2017) and 3rd in Southeast Asia, after only Singapore and Malaysia (Figure 1).

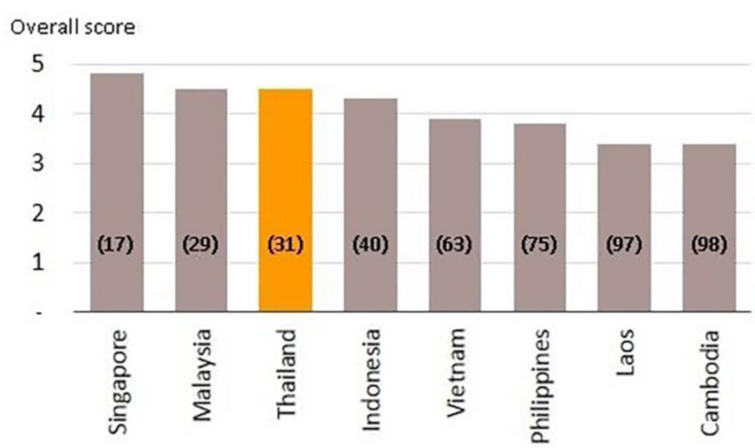


Figure 1 Travel & tourism competitiveness. **Source:** (Lunkam, 2021)

1.2 Aim

From above, the tourism and service industries make an important contribution to the country's economy. However, hotels, the most important segments of the tourism industry, are experiencing the negative impacts from the COVID-19 pandemic. COVID-19 is a major disruptor of the tourism and hospitality industries (Gossling, 2020). The service industries have had to develop more efficient approaches, innovate, and employ bold new strategies to attract the smaller customer base that has resulted from COVID-19 restrictions.

The pandemic is changing the behavior of tourists - they have become more sensitive toward the health and safety standards of firms and destinations (Wen et al., 2021). The pandemic has created an urgent need for contactless hospitality, which connects guests and hotel staff while reducing face-to-face interactions.

Contactless hospitality solutions include guest engagement, guest messaging, online ordering for room service, etc. In 2022, few hotels will be able to operate with the efficiency and care that is required in a post-COVID world without using technology. Physical distancing was introduced as a recommended social behavior of people (Nicola et al., 2020). As a result of tourists' fear of the virus and governments' actions to control it, tourism demand decreased (Dube et al., 2020).

The objective of this research is to use machine learning and artificial intelligence (AI) to help support the transformation of the hotel operation processes by developing a "contactless" software application as part of the "new normal" for the hotel guest and hotel owner. Self service and online hotel check-in solutions provide the ability to collect all necessary guest details automatically, with little to no contact from the reception desk. These details include passport scans, payment or deposit, E-signatures, arrival time, and more. This helps avoid unnecessary lines at reception, providing an improved arrival experience and first impression. Using online check-in allows hotels to receive all the information ahead of time, re-establish relationships with guests (often reduced due to online travel agency's), and offer additional services - it is the gateway to guest engagement and for digitizing the guest experience (Ganaban, 2022).

The software requirements are based on the needs of the hotel owner, as determined through interviews. The hotel owner wants to see, for example, a hotel management control panel that can determine the best possible room rate in real-time, as the market changes. The owner also would like to know customers' expectations about their experience with the hospitality services. Hospitality experience can be affected by various factors that are mostly based on the attitude and behavior of service staff such as being patient, welcoming guests, kindness, genuineness, and creating a comfortable atmosphere (Solnet et al., 2019). The hotel owner can be involved in the manual implementation of revenue management tasks. The system can effectively sift through the signals detected from market variables, discover patterns and anomalies, make predictions for guest arrivals, and calculate optimum prices in real-time.

As new data sets become available, the system can effectively gauge whether the information is important and if so, integrate it into current data parameters without the involvement of the hotel owner. As additional pertinent data are integrated into the existing parameters, the signals will change, making the pricing suggestions generated by the solution even more accurate. The hotel owner may receive too much, and unfiltered data, making it nearly impossible to process all the data effectively and determine an accurate price without the data analytics system. The machine-learning-based system also allows the implementation of dynamic rates based on specific variables chosen by the hotel owner. For example, a hotel could increase prices based on market demand signals obtained by analyzing vacation property demand, often an early indicator of future demand.

A limited view of market demand results in missed market demand signals, but when combined with other market data collected, dynamic pricing ensures that hotels are not ceding revenue unnecessarily, making the technology an important component of a proactive revenue management strategy. Algorithms can be designed to analyze sentiments in online guest chat and reviews and suggest operational improvements that can be made to improve guests' experiences. Over time, these sentiments can also be used as another basis for dynamic pricing; when guests are happy with their stay, the system will increase the room rate and if that changes, the room rates may decrease.

Machine learning and AI have influenced the development of various industries. The hotel industry can have a new experience in terms of satisfaction and service improvements for the guests using machine learning techniques (Yi, 2019). Machine learning technology can also be used to compute dynamic clusters of guests to create fluid segmentation in real-time. As consumer buying habits and/or booking patterns evolve, fluid segmentation ensures that the hotel continues to reach the right guests, at the right time and price, through the right channels. Due to the breadth of data available, only a machine-learning-based system can effectively leverage this information to dynamically maximize hotels' booking conversion rates and revenues.

Furthermore, according to Information Age (Ismail, 2016), "The hospitality industry has not always been at the forefront of high-tech innovation or implementation". While for other industries such as manufacturing, the adoption of AI is considered normal, the service sector is quite different. Service industries, including tourism, travel, and related industries, often call for personal contact between a provider and a client. The interplay in producing customer experience is at the core of the service business.

Until recently, most hotel bookings, transactions, and administrative tasks were handled manually. Revenue management - the process by which a hotel manager determines the best room rate at a given time, to maximize bookings and revenue - was a particularly difficult task. Hotel managers had to manually collect, review, and analyze numerous data sets each time the rate needed to be updated, and then calculate the ideal room rate based on those variables. Some hotel managers check online reviews to increase value by analyzing customer preferences and modifying services based on the available information (Noone et al., 2011). This was a very time-consuming task, which meant that hotel managers could not frequently and efficiently update rates. With the creation of online travel agencies, unparalleled quantities of data became available and the task of manually executing pricing decisions became impractical and outdated. Thus, hotel managers and policymakers are progressively looking to find travellers' preferences to improve their marketing policy and gain a smarter marketplace (De Pelamacker, 2018).

1.3 Statistical and Machine Learning Algorithms

This research used the following algorithms: regression algorithm such as linear regression, decision trees algorithm such as boosted decision trees, and clustering algorithm such as k-mean clustering on Microsoft Azure Machine Learning Studio.

1.3.1 Regression Algorithm

A regression algorithm normally is used to predict a response variable using independent variables such as predicting temperature of a city based on latitude, season, proximity to water bodies, cloudiness, population density, percent imperviousness, and extent of green space. The input variables may be numeric or categorical. However, the output of the regression analysis is numeric.

Linear regression is one of the oldest prediction technics in statistics. The goal of linear regression is to fit a linear model between the response and independent variables and use it to predict the outcome given a set of observed independent variables. The structure of a multiple linear regression model is (Barga et al., 2015)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \varepsilon$$

where

- Y_{is} the response variable that is trying to predict.
- X_1, X_2, X_3, ... are the independent variable used to predict the outcome.
- β_0 is a constant that is the intercept of the regression line, or the value. Of y when all of the predictor variables are 0.

- β_1, β_2, β_3, ... are the coefficients of the independent variables that are referred to the partial slopes of each variable.
- ε is the error or noise associated with the response variable that cannot be explained by the independent variables X_1, X_2, X_3, ...

1.3.2 Decision Trees Algorithm

Decision tree algorithms are hierarchical technics that work by splitting the dataset iteratively based on certain statistical criteria (the criteria depend on amount of the dataset). The goal of a decision tree is to maximize the variance across different nodes in the tree and minimize the variance within each node. Figure 2 shows a simple decision tree created with two splits of the data. The root node (Node 0) contains all the data in the dataset. The algorithm splits the data based on the criteria, creating three new nodes (Node 1, Node 2, Node 3). Using the same statistic, it splits the data again at Node 1, creating two more leaf nodes (Node 4, Node 5). The decision tree makes its prediction for each data row by traversing to the leaf nodes.

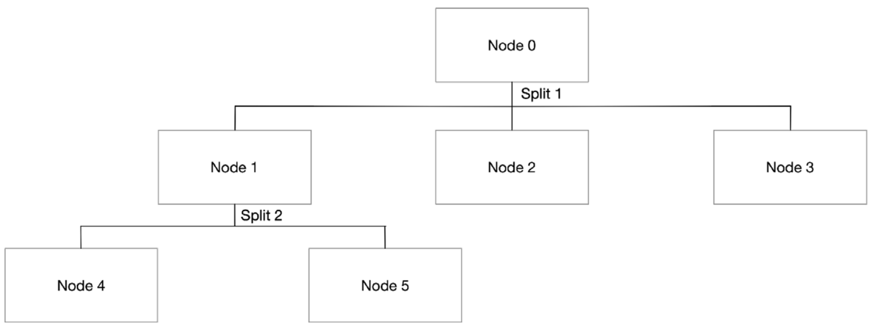


Figure 2 A simple decision tree with two data splits. Source: (Barga et al., 2015)

Boosted decision trees are a form of ensemble models. The models use several decision trees to produce superior predictors. Each of the individual decision trees can be weak predictors. However, when combined they can produce superior results (Barga et al., 2015).

1.3.3 Clustering Algorithm

The goal of clustering is to group similar objects together. Most existing cluster algorithms can be categorized as follows (Barga et al., 2015):

- Partitioning: Divide a dataset into k partitions of data. Each partition corresponds to a cluster.
- Hierarchical: Given a dataset, hierarchical clustering approaches start either bottom-up or top-down when constructing the clusters. In the bottom-up approach, the algorithm starts with each item in the dataset assigned to one cluster. As the algorithm moves up the hierarchy, it merges the individual clusters that are similar into bigger clusters. This continues until all of the clusters have been merged into one (root) of the hierarchy. In the top-down approach, the algorithm starts with all items in one cluster, and in each iteration, divides into smaller clusters.
- Density: Density-based algorithms grow clusters by considering the density (number of items) in the “neighbourhood” of each item. They often employed some form of a distance measure. This produces clusters that have regular shapes (e.g., spherical).

2. Literature Review

2.1 Role of AI in Service Industry Transformation

2.1.1 Key driving factors for AI

The key factors that drive AI (Figure 3) in the service industry for service providers to differentiate themselves with highly personalized services while keeping guest preferences in mind include: 1) Customer Experience: The process of understanding guest preferences and guest data 2) Operation Efficiency: AI is the medium to learn and act to improve efficiency in operation 3) Revenue Improvement: The introduction of data analytics with guest data help a service provider to cost-effectively manage revenue potential and 4) Sustainable Growth: Deep knowledge of individual customers enables the marketing team to precisely suggest products and services to the customer (Infosys, 2018).

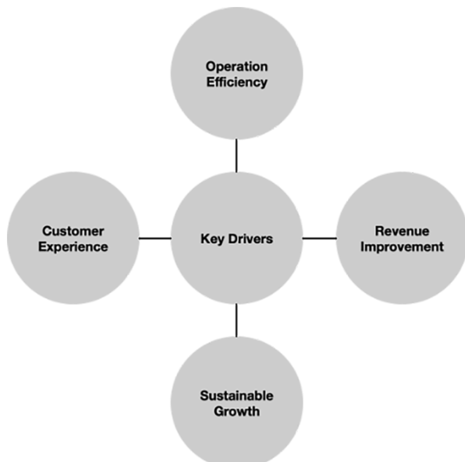


Figure 3 Key Drivers for AI. **Source**: Infosys (2018)

2.2 Data Analytics in the Service Industry

2.2.1 What is Data Analytics?

According to BMC Software (Kidd & Hornay, 2021), data analytics is the broad field of using data and tools to make business decisions and data analysis, a subset of data analytics, refers to specific actions.

The primary difference between analytics and analysis is a matter of scale, as data analytics is a broader term of which data analysis is a subcomponent. Data analysis refers to the process of examining, transforming, and arranging a given data set in specific ways to study its individual parts and extract useful information. Data analytics is an overarching science or discipline that encompasses the complete management of data. This not only includes analysis, but also data collection, organization, storage, and all the tools and techniques used.

It is the role of the data analyst to collect, analyse, and translate data into information that is accessible. By identifying trends and patterns, analysts help organizations make better business decisions. Their ability to describe, predict, and improve performance has placed them in increasingly high demand globally and across industries.

2.2.2 Hotel Data Types

From "Hotel Data Management: Solutions and Practices to Turn Information into a Valuable Asset" (AltexSoft, 2019), for the hospitality industry, the power of data helps decision-makers to solve the challenging domain-specific tasks including improving occupancy forecasting, setting competitive room prices, choosing the most profitable distribution channels, optimizing procurement operations, increasing guest loyalty, and identifying and targeting the most profitable guests (Figure4; Table1).

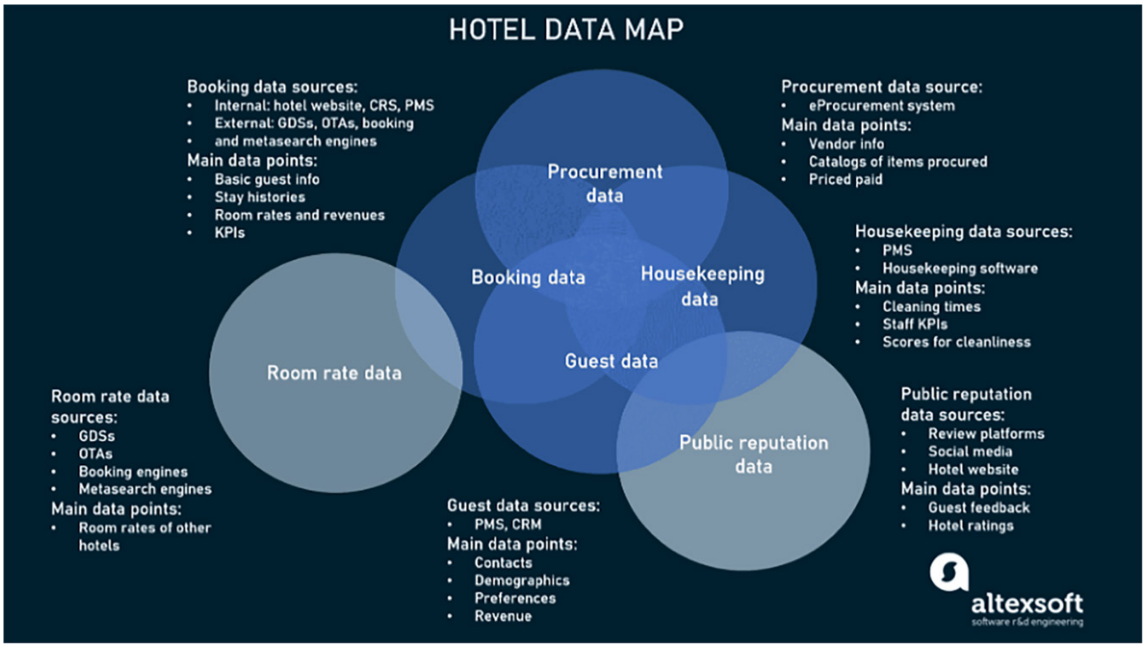


Figure 4 Hotel Data Map. **Source**: AltexSoft (2019)

Table 1. Hotel Data Types. **Source**: AltexSoft (2019)

Booking and Property Data	
Basic guest information	name, age, country of residence
Distribution channel	website, online travel agency, a metasearch engine, hotel chain, etc.
Lead time	advance time of booking
Length of stay (LOS)	duration of stay
Room price	a stay cost per night
Room history	name of guest and time per room
Key performance metrics (KPI)	average daily rate, occupancy rate, revenue per room

Service Data	
Hotel staff	number of staff working
Service time	service response times
Service KPI	performance indicators, quality of service
Service expense	expense
Service request	guest requests and complaints
Service score	guest scores for service

Room Rate Data	
Room rate	room price per day

Guest Data	
Contact information	phone number, physical and e-mail address
Demographics	age, marital status, number of children
Booking	reservation data, booking channel preferences, and past booking history
Reason	reasons for stay
Service	auxiliary services used
Food	food and beverage preferences
Cleaning	cleaning notes and housekeeping preferences
Revenue	revenue details
Payment	payment methods
Review	loyalty level (feedback)

Procurement Data	
Food	food and beverage
Furniture	furniture, fixtures a equipment
Disposable	disposables
Linen	linen and towels
Service	services
Uniform	uniforms
Print	printing

2.3 Hotel Data Analytics

The hotel industry possesses a wealth of data, encompassing a vast array of information. Unfortunately, most service providers fail to utilize this asset to its full potential. While many hoteliers collect loyalty data, they rarely delve into analytical processes to gain a deeper comprehension of their guests' behaviors, desires, and preferences. This limits their ability to develop a more understanding of their customers, tailor their services to meet their needs and preferences, and explore new way to attract additional clients.

Hotel Data Analytics is often used to segment guests according to booking trends, behaviour, and other factors to reveal their likelihood to respond to promotions and emerging travel trends. It is vitally important for hoteliers to be able to understand guest preferences (locations, activities, and room types), purchase behaviour (frequency, length of stay, time of year), and profit potential to increase brand loyalty and wallet share of their most valuable guests.

Regrettably, financial resources are frequently allocated towards general advertising campaigns that fail to focus on specific guests or market segments with tailored promotions that are more likely to elicit a positive response. Consequently, guests may perceive that the hotel lacks concern for their individual needs, or that its services are not tailored to meet their specific requirements. Such an experience could motivate guests to switch to a competing hotel with a more personalized approach.

To truly leverage the power of analytics, hospitality organizations must distinguish between reactive and proactive decision-making. While historical data can provide reports, drilldowns, or alerts to monitor customer trends, advanced analytics can help hotels identify the root causes of these trends, forecast future outcomes, and determine the best course of action, while considering all operational constraints. By utilizing such advanced analytics, hotels can achieve a deeper understanding of their patrons, and thereby create a significant impact on their operations (Dragosavac, n.d.).

2.3.1 Customer Segmentation

Initially, the hotel's task is to identify distinctive cluster groups, followed by performing a value segmentation analysis for each group separately. Later, the hotelier can review their present business customer base and create unique clusters of business customers. For instance, they may recognize that some business customers only use the hotel for brief overnight stays, whereas others are there for longer-term events conducted at the hotel. It is feasible to break down these groups even further by examining other behavioural and relationship aspects (Dragosavac, n.d.).

2.3.2 Customer Profiling

To profile customers, a comprehensive examination of guest demographics and lifestyle traits is conducted. Characteristics such as income, family status, age, and interests in sports and culture, if available, can be added to shape a model of guests. This technique can be leveraged to generate an email roster for focused marketing campaigns aimed at both existing and potential customers. Identifying prospect profiles can be particularly advantageous in pinpointing those who are more likely to react positively to marketing and promotional proposals. Profiling can also assist in determining which market segments are the most productive and profitable (Dragosavac, n.d.).

*2.3.3 Menu Engineer*ing

A review of the sales of menu items and their contribution margins can aid in sustaining a growth of restaurant business. While menu engineering addresses menu content choices, data mining can generate reports that show the preferred menu item selections by customer segment, which can serve as a foundation for enhancing operations. For example, Applebee's has been known to utilize data mining to determine ingredient restocking amounts, leveraging a menu optimization quadrant analysis that summarizes the sales of menu items. Such analyses help the hotel decide which menu items to promote (Dragosavac, n.d.).

2.3.4 Productivity Indexing

Through a correlation of the time of order entry with the time of settlement, hoteliers can make more precise estimates of production and service times. These data offer an understanding of the average service time regarding customer turnover, along with wait time statistics. Although productivity data can be challenging to gather, this analysis provides concrete information to aid management in refining operations (Dragosavac, n.d.).

2.3.5 Customer Associations and Sequencing

Advanced big-data analytics can reveal correlations between seemingly disconnected events. For example, if a guest orders the restaurant's signature dish, they are likely to also order a small antipasto salad and a glass of Chardonnay. These paired relationships can form the basis for bundling menu items into a cohesive meal, simplifying the ordering process while ensuring customer contentment. Menu design can be adjusted to showcase such combinations as distinctive opportunities for customers. Data associations are often considered a means of persuading customers to spend more than they intended or to upsell (Dragosavac, n.d.).

2.3.6 Forecasting

Forecasting empowers restaurants to proactively exceed their clients' requirements by optimizing staffing, procurement, preparation, and menu design. Customers perceive their stay at a hotel as an experience rather than just a visit, and various activities, such as fine dining, nightly entertainment, spas, corporate seminars, or meetings, contribute to enhancing their experience. However, not all activities are equally appealing to all clients. In this regard, analytics can play a pivotal role in enabling hotels to comprehend the diverse needs of their clients more effectively (Dragosavac, n.d.)

2.3.7 Energy Consumption

Analytics can be leveraged for internal operations in the hotel industry as well. Energy consumption constitutes a significant portion of a typical hotel's utility costs, accounting for 60 to 70 percent. However, it is possible to manage costs efficiently while maintaining guest comfort by optimizing energy usage. Managers can leverage smart data to create energy profiles for their hotels. Modern software solutions can collate data from multiple sources, including weather forecasts, electricity rates, and a building's energy consumption, to develop a comprehensive "building energy profile." With the aid of a cloud-based, predictive analytics algorithm, the software can optimize whether the power supply comes from the grid or an onsite battery module (Dragosavac, n.d.).

2.3.8 The Right Room at the Right Rate

Although yield management has been a practice in the hotel industry for some time, big data offers hotels the opportunity to take revenue management to the next level by providing truly personalized prices and rooms to guests. Industry studies suggest that Marriott, a leading hotel chain, has leveraged big data analytics to predict the ideal price of its rooms to fill its hotels. By utilizing advanced revenue management algorithms that can process data at a faster pace, Marriott has been able to combine various data sets to derive insights that are accessible to all levels, thereby enhancing decision-making capabilities (Dragosavac, n.d.).

3. Methodology

An in-depth study was conducted through the cooperation of two boutique hotels located in the center of Bangkok and that began operation two years ago. Most of their guests are from America, Europe, and China. The researcher studied target users, including their needs and pain points. In this place, the target user is the owner, and the hotel manager interviews the target user by asking questions about the hotel process to get the user's requirements.

Based on the Microsoft Azure Machine Learning Studio Documentation (Gilley, 2022), three algorithms were selected for each approach. The Logistic Regression and Boosted Decision Tree were used to predict the prospective customer, and the K-Mean Clustering was used for a type of customer segmentation. Similar algorithms were selected to train the model because the results can be compared between each algorithm.

3.1 Data acquisition and preparation

Booking data for the study were obtained from each hotel 3946 records available from Hotel A and 3948 records from Hotel B. The period of record was April 2016 to September 2021 (Figure 5). Each dataset contained the following: 1. Hotel Name 2. Book Number 3. Book By 4. Guest Name 5. Book On 6. Check In 7. Lead Time 8. Check Out 9. Stay 10. Status 11. Channel 12. Price

	A	B	C	D	E	F	G	H	I	J	K	
1	Hotel Name	Book Number	Book By	Guest Name	Book On	Check In	Lead Time	Check Out	Stay	Status	Channel	Price
2	White Ivory	450058371	ABC DEF	ABC DEF	23/4/2016	23/4/2016	0	24/4/2016	1	Canceled	OTA	
3	White Ivory	160029890	ABC DEF	ABC DEF	9/5/2016	10/5/2016	1	11/5/2016	1	OK	OTA	
4	White Ivory	547873297	ABC DEF	ABC DEF	9/5/2016	11/5/2016	2	12/5/2016	1	OK	OTA	
5	White Ivory	855203406	ABC DEF	ABC DEF	24/3/2016	20/5/2016	26	21/5/2016	1	OK	OTA	
6	White Ivory	456649527	ABC DEF	ABC DEF	11/4/2016	21/5/2016	10	22/5/2016	1	Canceled	OTA	
7	White Ivory	243347585	ABC DEF	ABC DEF	25/3/2016	22/5/2016	27	23/5/2016	1	OK	OTA	
8	White Ivory	505182861	ABC DEF	ABC DEF	22/5/2016	2/6/2016	11	3/6/2016	1	OK	OTA	
9	White Ivory	527806824	ABC DEF	ABC DEF	21/5/2016	2/6/2016	12	3/6/2016	1	Canceled	OTA	
10	White Ivory	991623938	ABC DEF	ABC DEF	22/5/2016	3/6/2016	12	5/6/2016	2	OK	OTA	
11	White Ivory	257882765	ABC DEF	ABC DEF	19/5/2016	29/6/2016	10	14/7/2016	15	OK	OTA	
12	White Ivory	234331915	ABC DEF	ABC DEF	25/5/2016	18/7/2016	23	24/7/2016	6	Canceled	OTA	
13	White Ivory	784277594	ABC DEF	ABC DEF	22/6/2016	25/7/2016	3	26/7/2016	1	OK	OTA	
14	White Ivory	981176830	ABC DEF	ABC DEF	10/7/2016	26/7/2016	16	27/7/2016	1	OK	OTA	
15	White Ivory	693511725	ABC DEF	ABC DEF	8/7/2016	4/8/2016	27	9/8/2016	5	Canceled	OTA	
16	White Ivory	642819934	ABC DEF	ABC DEF	5/7/2016	5/8/2016	0	9/8/2016	4	Canceled	OTA	
17	White Ivory	767922529	ABC DEF	ABC DEF	15/6/2016	10/8/2016	26	15/8/2016	5	Canceled	OTA	
18	White Ivory	987801344	ABC DEF	ABC DEF	10/8/2016	10/8/2016	0	12/8/2016	2	OK	OTA	
19	White Ivory	595018378	ABC DEF	ABC DEF	2/8/2016	11/8/2016	9	8/9/2016	28	Canceled	OTA	
20	White Ivory	314062568	ABC DEF	ABC DEF	11/8/2016	12/8/2016	1	13/8/2016	1	OK	OTA	

Figure 5 Hotel Booking Data.

3.2 K-Mean Clustering

In general, clustering uses iterative techniques to group cases in a dataset into clusters that possess similar characteristics. These groupings are useful for exploring data, identifying anomalies in the data, and eventually for making predictions. Clustering models can also help to identify relationships in a dataset that might not be logically derived by browsing or simple observation. For these reasons, clustering is often used in the early phases of machine learning tasks, to explore the data and discover unexpected correlations. To configure a clustering model using the K-means method, a target number k must be specified that indicates the number of centroids in the model. The centroid is a point that is representative of each cluster. The K-means algorithm assigns each incoming data point to one of the clusters by minimizing the within-cluster sum of squares (Gilley, 2022).

3.3 Logistic Regression

Logistic regression is a statistical method that is commonly employed in classification tasks to estimate the probability of a particular outcome. This technique involves using a logistic function to model the data and determine the likelihood of an event occurring (Gilley, 2022).

3.4 Boosted Decision Tree

Boosting is an ensemble modeling method that is frequently combined with bagging, random forests, and other techniques. Azure Machine Learning features an efficient implementation of gradient boosting algorithm, which utilizes boosted decision trees. Gradient boosting is a commonly used machine learning method for regression tasks, in which regression trees are constructed incrementally using a predefined loss function to measure and correct errors in each iteration. The final prediction model is a combination of weaker models, leading to improved overall performance (Gilley, 2022).

3.5 Train the Model

In Azure Machine Learning Designer Studio, Figure 6 shows the regression model is trained to predict the prospective customer. The "hotel-data" were the hotel booking data, "hotel-data.csv", the CSV file was uploaded to the dataset container and used on the first module. The module "Clean Missing Data" was used to delete rows of missing data. The module "Split Data" was used to split the dataset for training and testing 60 percent and 40 percent respectively. The algorithms "Two-Class Logistic Regression" and "Two-Class Boosted Decision Tree" were selected to train the model at the same time, and the module "Train Model" was involved in training with the dataset, whereas the module "Score Model" presented the result of the model testing. The prediction model performance was done by the module "Evaluate Model".

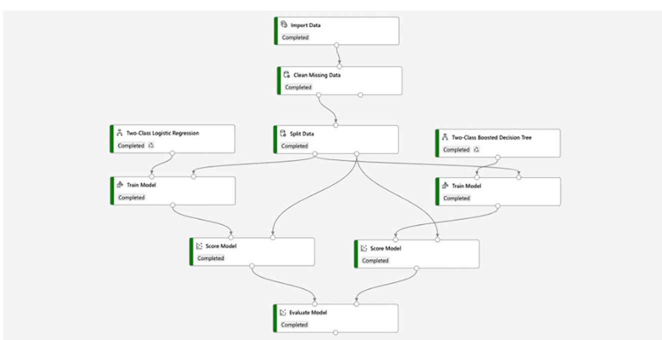


Figure 6 Train the Regression Model.

Figure 7 shows the clustering model is trained to predict customer segmentation. The CSV file, “hotel-data” was uploaded to the dataset container and used on the first module. The module “Split Data” was used to split the dataset for training and testing 60 percent and 40 percent respectively. The algorithm “K-Mean Clustering” was selected to train the model with the number of centroids set as 2 for groups of customers, and the module “Train Clustering Model” was involved in training with the dataset. The module “Assign Data to Clusters” was used to assign a group number to each row of the dataset.

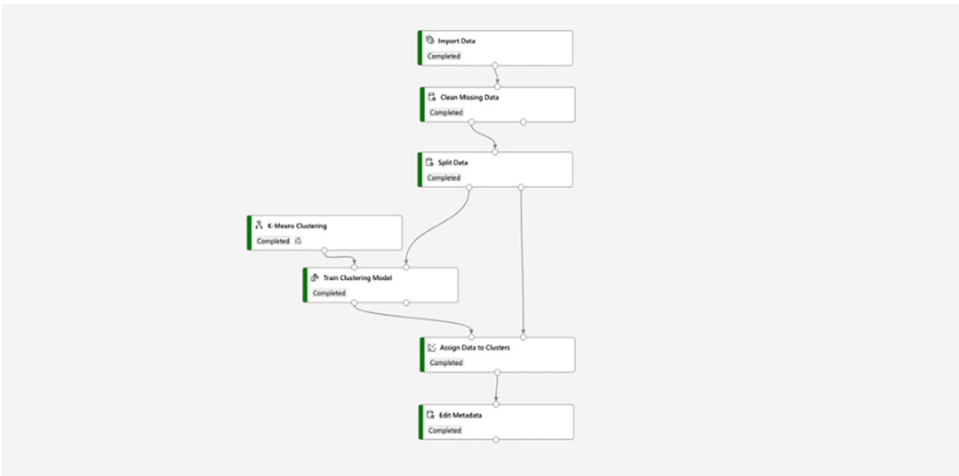


Figure 7 Train the Clustering Model.

4. Result

4.1 Score the Model

Figure 8 shows, the results from the Two-Class Logistic Regression model testing on the module “Score Model”. From the “Score Model Result Visualization”, the “Scored Labels” column was added and indicates the prediction results of the newly trained model. The prediction value is nearly 1 (the value is between 0 to 1) on the “Score Labels” column, so it notes that the model is accuracy is satisfactory.

Figure 8 Two-Class Logistic Regression Score Model Result.

Results from the regression model testing on the module “Score Model” are presented in Figure 9. From the “Score Model Result Visualization”, the “Scored Labels” column was added. Numerical results of the prediction value on the “Score Labels” column and the actual value is nearly 1 (the value is 0.9787), which indicates that this model is accurate.

Rows 1,973 Columns 14

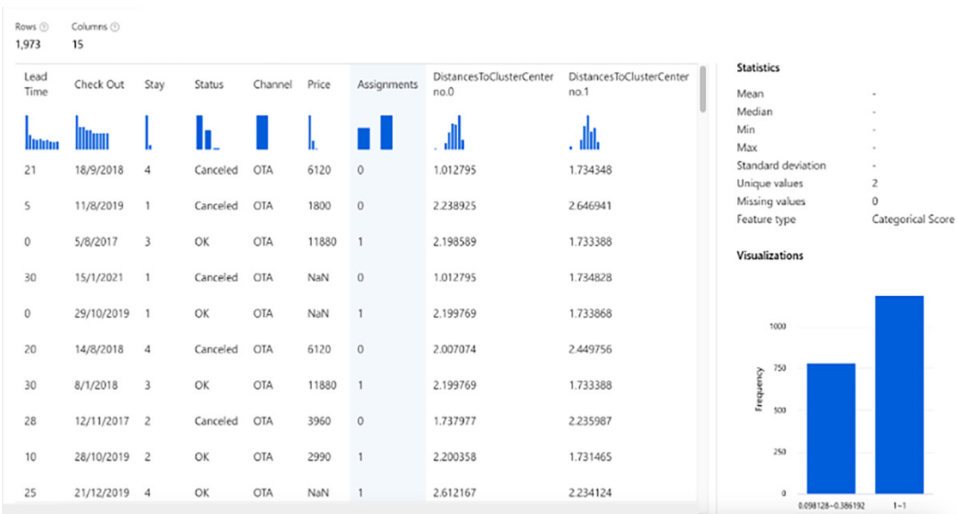
Guest Name	Book On	Check In	Lead Time	Check Out	Stay	Status	Channel	Price	Scored Labels	Scored Probabilities
	2018-06-19 00:00:00	2018-07-19 00:00:00	0	2018-07-23 00:00:00	4	OK	OTA	6336	White Ivory	1
	2019-02-03 00:00:00	2019-02-12 00:00:00	9	2019-02-13 00:00:00	1	OK	OTA	0	White Ivory	1
	2016-11-26 00:00:00	2017-02-24 00:00:00	29	2017-02-26 00:00:00	2	OK	OTA	4400	White Ivory	1
	2018-09-18 00:00:00	2019-04-12 00:00:00	25	2019-04-18 00:00:00	6	Canceled	OTA	7293	White Ivory	1
	2019-09-15 00:00:00	2020-04-10 00:00:00	26	2020-04-15 00:00:00	5	Canceled	OTA	9639	White Ivory	1
	2019-12-28 00:00:00	2020-01-17 00:00:00	20	2020-01-19 00:00:00	2	OK	OTA	0	White Ivory	1
	2019-11-26	2019-11-27	1	2019-11-28	1	OK	OTA	1000	White Ivory	1

Statistics

Mean	0.9787
Median	1
Min	0
Max	1
Standard deviation	0.1444
Unique values	1194
Missing values	0
Feature type	Numeric Score

Figure 9 Two-Class Boosted Decision Tree Score Model Result.

As shown in Figure 10, results from the clustering model testing on the module “Assign Data to Cluster”, the “Assignments” column was added. The “Assignments” histogram shows that the hotel data have effectively clustered into two discrete groups numbered 0 to 1 (the unique values is 2).



Rows 1,973 Columns 15

Lead Time	Check Out	Stay	Status	Channel	Price	Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1
21	18/9/2018	4	Canceled	OTA	6120	0	1.012795	1.734348
5	11/8/2019	1	Canceled	OTA	1800	0	2.238925	2.646941
0	5/8/2017	3	OK	OTA	11880	1	2.198589	1.733388
30	15/1/2021	1	Canceled	OTA	NaN	0	1.012795	1.734828
0	29/10/2019	1	OK	OTA	NaN	1	2.199769	1.733868
20	14/8/2018	4	Canceled	OTA	6120	0	2.007074	2.449756
30	8/1/2018	3	OK	OTA	11880	1	2.199769	1.733388
28	12/11/2017	2	Canceled	OTA	3960	0	1.737977	2.235987
10	28/10/2019	2	OK	OTA	2990	1	2.200358	1.731465
25	21/12/2019	4	OK	OTA	NaN	1	2.612167	2.234124

Statistics

Mean	-
Median	-
Min	-
Max	-
Standard deviation	-
Unique values	2
Missing values	0
Feature type	Categorical Score

Figure 10 Clustering Score Model Result.

4.2 Evaluate the Model

The “Evaluate Model” module was used to measure the performance of the trained model. This module takes two datasets as inputs. The first is a scored dataset from a tested model. The second is an optional dataset for comparison. The result from an optional model should not be very different from a tested model.

4.3 Model Metrics

A set of curves and metrics are produced to view the results of a scored model or compare the results of two scored models. The results are in the following formats (Berns, 2015):

- Receiver Operating Curve (ROC) which displays the fraction of true positives out of the total actual positives. It contrasts this with the fraction of false positives out of the total negatives, at various threshold settings. The diagonal line represents above 50 percent accuracy in the prediction that can be used as a benchmark of improvement. The higher and further to the left, the more accurate the model.
- Precision-recall Curve represents the fraction of retrieved instances that are relevant, whereas recall represents the fraction of relevant instances that are retrieved.
- Lift Curve is a variation of the ROC curve. It measures the fraction of true positives, in relation to the target response probability.

The metrics returned for the models are designed to estimate the amount of error in model estimates. A model is considered to fit the data well if the difference between observed and predicted values is small. However, looking at the pattern of the residuals (the difference between the predicted point and its corresponding actual value) can tell a lot about potential bias in the model. The following metrics are reported for evaluating the models (Gilley, 2022).

- The Mean Absolute Error (MAE) assesses the proximity of the predictions to the actual outcomes, where a lower score is indicative of better performance.
- The Root Mean Squared Error (RMSE) calculates a singular value that summarizes the model's error. By squaring the difference between predicted and actual values, this metric ignores the distinction between over-prediction and under-prediction.
- Similarly, the Relative Squared Error (RSE) normalizes the total squared error of predicted values by dividing it by the total squared error of actual values.

4.4 Model Performance

Figure 11 illustrates, the "Evaluate Model Result Visualization" representing the "Left port" and the "Right port" which are the results from the "Evaluate Module". The "Left port" is the result from training by the "Two-class Logistic Regression" algorithm, and the "Right port" is the result from training by the "Two-class Decision Tree" algorithm. Both ROC curves are in the highest position.

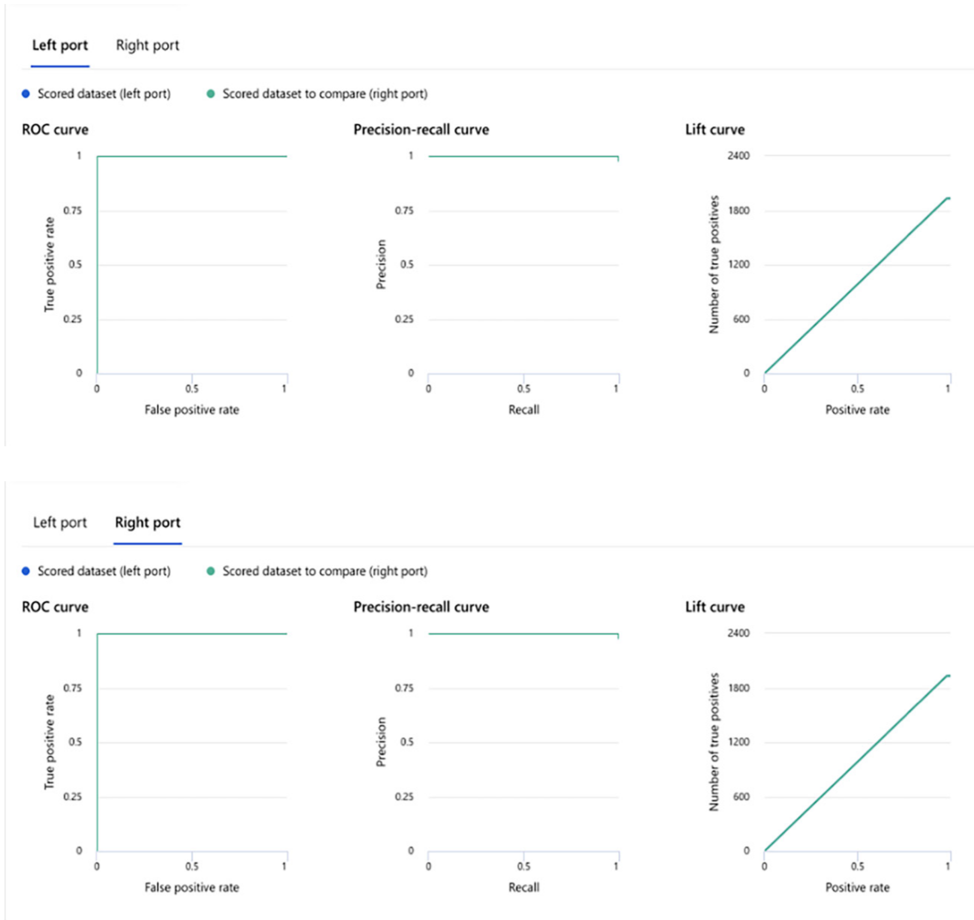
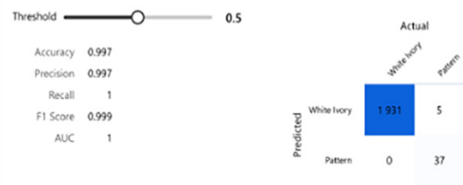
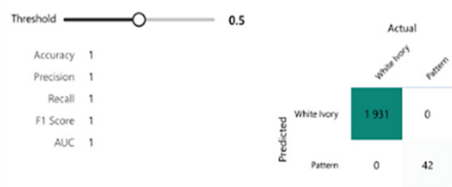


Figure 11 The Model Performance Evaluation Comparison.

As shown in Figure 12, the “Evaluate Model Result Visualization” represents the matrix value from each algorithm, and shows Accuracy, F1 Score, Precision, and Recall.



Score bin ↓	Positive exam...	Negative exam...	Fraction above thresh...	Accuracy	F1 Score	Precisi...	Recall	Negative precisi...	Negative recall	Cumulative AUC
(0.900,1.000]	1921	0	0.974	0.995	0.997	1.000	0.995	0.808	1.000	0.000
(0.800,0.900]	9	0	0.978	0.999	1.000	1.000	0.999	0.977	1.000	0.000
(0.700,0.800]	1	1	0.979	0.999	1.000	0.999	1.000	1.000	0.976	0.024
(0.600,0.700]	0	1	0.980	0.999	0.999	0.999	1.000	1.000	0.952	0.048
(0.500,0.600]	0	3	0.981	0.997	0.999	0.997	1.000	1.000	0.881	0.119
(0.400,0.500]	0	2	0.982	0.996	0.998	0.996	1.000	1.000	0.833	0.167



Score bin ↓	Positive exam...	Negative exam...	Fraction above thresh...	Accuracy	F1 Score	Precisi...	Recall	Negative precisi...	Negative recall	Cumulative AUC
(0.900,1.000]	1931	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.800,0.900]	0	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.700,0.800]	0	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.600,0.700]	0	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.500,0.600]	0	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.400,0.500]	0	0	0.979	1.000	1.000	1.000	1.000	1.000	1.000	0.000

Figure 12 The Model Matrix Comparison.

5. Discussion

The results generated by AI models can provide valuable insights for study hotels seeking to improve their operations. For example, an AI model may analyze guest booking data to identify patterns in customer behavior and preferences. This information can then be used by hotels to develop targeted marketing campaigns and personalized guest experiences. AI models can also be used to optimize hotel pricing and revenue management strategies. By analyzing historical booking data and market trends, AI algorithms can generate predictions about future demand and recommend pricing strategies to maximize revenue.

In addition, AI models can help hotels to streamline their operations by identifying inefficiencies and bottlenecks in processes such as housekeeping and maintenance. By automating certain tasks and providing real-time insights, AI can help to reduce costs, improve guest satisfaction, and increase overall efficiency. Overall, AI models can provide powerful tools for study hotels to improve their operations, enhance guest experiences, and increase profitability.

Technological advances are causing fundamental disruption in the tourism industry by empowering tourism actors to form new markets, shape new services, and manage their businesses more effectively (Law et al., 2014). Technologies such as Machine Learning can open a whole world of possibilities for data analysis to attract new customers, retain old customers, and make the next stay better. COVID-19 was a striking health crisis that sped up the adoption of AI and triggered studies discussing the implementation of safety issues. However, there still are few studies that examine the perception of service industries toward AI (Quanovo Data Technologies, 2017). The reliance on technologies through interoperability and interconnectivity of all network partners increasingly enabled hospitality organisations to develop their competitiveness through a better understanding of customers and market conditions and develop their decision-making process (Buhalis & Foerste, 2015)).

6. Conclusion and Recommendation

The purpose of this study was to create machine learning models for the boutique hotel industry. The difficulty of building models was lack of data. It was only possible to use hotel booking data for model creating, testing, and training with the Azure Machine Learning tool, and the hotel only had that kind of data for data analytics. For the prediction results from the tool, there are two approaches. Firstly, to train a model for finding the potential customer. Secondly, to train a model for grouping customers. The outcome from the two approaches is fairly satisfactory. However, more data are required to create the models that were discussed above, for example, the Menu Engineering Model needs food and beverage buying data, or the Productivity Indexing Model needs service time data, and etc. Thus, the next stage of the research should include gathering the necessary data that can be analyzed to better address the hotel owner's requirements for a contactless software solution.

Author Contributions

Conceptualization, S.N and C.B; methodology, S.N and C.B; Validation, S.N and C.B; formal analysis, S.N and C.B; investigation, S.N and C.B; resources, S.N and C.B; data Curation, S.N; writing - original draft, S.N; writing - review & editing, S.N; visualization, S.N; supervision, S.N and C.B; project administration, S.N and C.B; funding acquisition, S.N and C.B. All authors have read and agreed to the published version of the manuscript.

References

AltexSoft. (2019). *Hotel data management: Solutions and practices to turn information into a valuable asset.* Retrieved January 15, 2022, from https://www.altexsoft.com/blog/hotel-data-management-best-practices

Barga, S. R., Fontama, V., & Tok, W.-H. (2015). *Predictive analytics with microsoft machine learning: Build and deploy actionable solutions in minute*s. Apress.

Berns, J. (2015). *Azure machine learning*. Microsoft Press.

Buhalis, D., & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. *Journal of Destination Marketing & Management, 4*(3), 151–161. https://doi.org/10.1016/j.jdmm.2015.04.001

De Pelamacker, P., van Tilburg, S., & Holthof, C. (2018). Digital marketing strategies, online reviews and hotel performance. *International Journal of Hospitality Management, 72*, 47–55. https://doi.org/10.1016/j.ijhm.2018.01.003

Dragosavac, G. (n.d.). *Big data analytics : Analytics in hotel industry*. Retrieved January 30, 2022, from http://www.bigdatanalysis.com/big-data-analytics-in-hotel-industry

Dube, K., Nhamo, G., & Chikodzi, D. (2020). COVID-19 cripples global restaurant and hospitality industry. *Current Issues in Tourism, 24*(11), 1487-1490. https://doi.org/10.1080/13683500.2020.1773416

Ganaban, R. (2022). *Use next-gen tech for a contactless guest experience*. HotelTechReport. Retrieved February, 28, 2022 from https://hoteltechreport.com/news/next-gen-tech-contactless-guest-experience

Gossling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: A rapid assessment of COVID-19. *Journal of Sustainable Tourism, 29*(1), 1-20. https://doi.org/10.1080/09669582.2020.1758708

Gilley, S. (2022). Azure Machine Learning documentation. Microsoft Corporation. Retrieved April 30, 2022, from https://docs.microsoft.com/en-us/azure/machine-learning

Infosys. (2018). *Role of AI in travel and hospitality industry*. Retrieved March 15, 2022, from https://www.infosys.com/industries/travel-hospitality/documents/ai-travel-hospitality.pdf

Ismail, N. (2016). *Machine learning and AI technology in the hospitality industry*. Information Age. Retrieved March 30, 2022, from https://www.information-age.com/machine-learning-hospitality-2982

Kidd, C., & Hornay. R. (2021). *Data analytics vs data analysis: What's the difference?*. BMC. Retrieved February 15, 2022, from https://www.bmc.com/blogs/data-analytics-vs-data-analysis

Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management, 25*(5), 727-750. https://doi.org/10.1108/IJCHM-08-2013-0367

Lunkam, P. (2021). *Nǣo nōm thurakit / ʻutsāhakam pī 2564-2566 : thurakit rōngrǣm* [Industry Outlook 2021-2023: Hotel Industry]. Krungsri Research. Retrieved April 15, 2022, from https://www.krungsri.com/th/research/industry/industry-outlook/services/hotels/io/io-hotel-21

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. *International Journal of Surgery, 78*, 185–193. https://doi.org/10.1016/j.ijsu.2020.04.018

Noone, B. M., McGuire, K. A., Rohlfs, K. V. (2011). Social media meets hotel revenue management: Opportunities, issues and unanswered questions. *Journal of Revenue and Pricing Management, 10*, 293–305.

Quanovo Data Technologies. (2017). *Machine learning technologies for the hospitality industry*. Retrieved May 15, 2022, from https://www.quanovo.com/Content/Whitepapers/machine-learning-technologies-for-the-hospitality-industry.pdf

Rob L., Dimitrios B., Cihan C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management. 26*(5) 727-750, 2014.

Solnet, D., Subramony, M., Ford, R. C., Golubovskaya, M., Kang, H. J., & Hancer, M. (2019). Leveraging human touch in service interactions: Lessons from hospitality. *Journal of Service Management, 30*(3), 392-409. https://doi.org/10.1108/JOSM-12-2018-0380

Wen, J., Kozak, M., Yang, S., & Liu, F. (2021). COVID-19: Potential effects on Chinese citizens' lifestyle and travel. *Tourism Review, 76*(1), 74-87. https://doi.org/10.1108/TR-03-2020-0110

Yi, R. (2019). Research on development status and trend of unmanned wisdom hotel system based on artificial intelligence service -take Bangwei Company as an example. *Finance and Market, 4*(1) 36-43.