ORIGINAL PAPER

Time changes of customer behavior on accommodation reservation: a case study of Japan

Koichi Ito[1] · Shunsuke Kanemitsu[2] · Ryusuke Kimura[3] · Ryosuke Omori[1]

Abstract

The forecasting of demand or cancellations is highly important for efficient revenue management in the hotel industry. Previous studies have mainly focused on the accuracy of the prediction of reservation number or cancellation rate on a specific accommodation or hotel chain; therefore, the application of the prediction to different accommodations or under the behavioral change of customers in response to natural or human events is difficult without the re-estimation of the prediction model. Information of the customer behavioral trend on the accommodation reservations is necessary for the construction of a general forecasting model. In this study, we focus on one of the general trends of customer behavior, that is, the reservation timing and the time changes of the cancellation probability using the big data of the reservation records provided by an online trip agency in Japan. We showed that the reservation timing and cancellation probability can be decomposed by five and six exponential functions of the days until the stay and the days from the reservations. We also showed that the significant factors influencing the time changing patterns are the guest numbers per room for both reservation and cancellation, composition of guests in terms of the number and gender of guests, and the stay length for reservation. These findings imply that the customer behavior during accommodation reservation could be categorized into multiple motivational factors toward reservations or cancellations. Our results contribute to the construction of a general forecasting model on the accommodation reservations.

✉ Ryosuke Omori
  omori@czc.hokudai.ac.jp

1 Division of Bioinformatics, International Institute for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido 001-0020, Japan

2 Data Solution Unit 2(Marriage and Family/Automobile Business/Travel), Data Management and Planning Office, Product Development Management Office, Recruit Co., Ltd, Chiyoda-Ku, Tokyo 100-6640, Japan

3 SaaS Data Solution Unit, Data Management and Planning Office, Product Development Management Office, Recruit Co., Ltd, Chiyoda-Ku, Tokyo 100-6640, Japan

1 Introduction

Forecasting is crucial for efficient revenue management [8, 28]. For the implementation of the pricing and capacity control, accurate prediction of the future demand, cancellation probabilities, or no-show rate is necessary. Various studies on forecasting have been conducted in the airline industry [5, 15]; nonetheless, such studies have recently been extended to other various industries, such as healthcare, rail industries, restaurants, convention centers, car rental agencies, cruise lines, golf courses, internet service, and the hotel industry [4, 8, 19, 30]

Until now, forecasting of various measurements of reservation has been examined [28, 29]. One of the targets of forecasting has been the final number of reservations, that is, the number of reservations not cancelled on the day of the reservation. Conventionally, the final number of the reservations has been predicted from the historically aggregated number of reservations [1, 15, 16]. Furthermore, the algorithms recently suggested tend to use the details of the reservations [30] that is also known as the passenger name record (PNR) format, a detailed booking data format including the detail of the customers standardized in the airline industry [12]. Although these studies have mainly focused on the previous total demand that can be capped by the total number of rooms in the focal accommodations, the estimation of the true demand was also attempted [15, 18, 25]. Another important target of forecasting is the cancellation probability and no-show rate. Particularly, some studies investigated the factors related to the cancellation probabilities based on the PNR format [3, 9, 23]. These previous studies mainly focused on the forecast accuracy because of its influence on business profit; indeed, a 10% improvement in the forecasting accuracy on high demand flights can cause a 1.5–3.0% increase in revenue [15].

However, the studies focusing on the improvement of the accuracy have some limitations in terms of the application of the predictions, particularly on the forecasting in the hotel industry. First, most previous studies have focused only on a specific accommodation or hotel chains [8]. Because the estimation only from a few accommodations could be biased owing to the unique features of the focal accommodations such as the capacity or the local events around the accommodations, application of the outcomes to other accommodations without the re-estimation and adjustment of the forecasts is hindered. Second, the forecasting based on the past data could not be applicable when the customers behaviors vary [28, 31]. Particularly, the customer behavior varies drastically in response to various types of crisis [2, 7, 21], such as natural or human disaster, political instability, economic recession, or the spread of disease [10, 11, 17, 20, 26]. The importance of the understanding of such behavioral changes of customers has been often suggested [21], although a few studies have investigated forecasting under such situations.

These limitations are caused by the lack of basic information regarding the common behavioral patterns of customers in accommodation reservations. The net demand,

cancellation probabilities, or booking curve are the aggregation of the decision-making of customers; therefore, the understanding of the decision-making patterns of customers at each accommodation reservation is important for examining the common trends that can be widely applied to various accommodations and situations. However, only few studies have focused on the common behavioral patterns of customers in individual accommodation reservation. For representing the time change of the reserved seat numbers in airlines, Lee [15] proposed a stochastic model that explicitly assumes the occurrence of the new reservation and the cancellations that are represented as Poisson and binomial processes, respectively. However, Lee [15] focuses on providing the method and therefore, the observed customer behavioral patterns during reservation has been discussed to some extent. Modelling the choice of consumers based on the feature (such as, flight time, price, and refund) has been also conducted [6, 27]. The analyses of the reservation or cancellation behavior linked with the customer personal details based on the PNR format [3, 9, 23] might enable the understanding of the background of customer decision making; nonetheless, the main aim in these studies were the improvement of the prediction accuracy and so there little implications.

For investigating the behavioral trends of general customers on accommodation reservations, in this study, we focused on big data, that is, the reservation records on the online travel agency (OTA). It contains extensive amounts of records for various types of accommodations, and therefore, such data are more suitable for the investigation of the general patterns or trends of customer behaviors compared with the detailed but narrow data, such as the dataset focusing only on specific hotel or hotel chain. Using the reservation records on OTA, in this study, we have focused on the time changes of the customer behavior on the accommodation reservations. In general, the reservation frequency or the cancellation rate tends to vary with the proximity of the date of the stay. This suggests that the motivation behind making new reservations or cancellations also vary with time. In other words, by investigating the general trends in the reservation timing or the time changes of cancellation probabilities, we can reveal the motivations triggering customer behavior during accommodation reservations.

Our aim in this study is to reveal the time changes of the reservation behavior of the consumers, that is, how long prior to the stay consumers make accommodation reservations, or how often and when the reservations are cancelled. Based on the PNR format accommodation reservation dataset obtained from an OTA in Japan, we decomposed the complicated reservation timing and the cancellation probabilities depending on the days until the stay in the accommodation into several elemental trends by fitting exponential functions. Moreover, using the reservation categories contained in the dataset, we also investigated the significant factors that influence the time changes of the customer behavior.

2 Materials and methods

2.1 Data

The PNR format accommodation reservation dataset was obtained from "jalan. net" (https://www.jalan.net), one of the largest online travel agents in Japan

[14]. All reservation records of the accommodations located in four prefectures, Miyagi, Aichi, Osaka, and Fukuoka, from December 28, 2015 to December 29, 2019 were included in the analysis. The number of accommodations located in four prefectures (Miyagi, Aichi, Osaka, and Fukuoka) provided at jalan.net are 318, 543, 629, and 575 (counted on "jalan.net" website on February 8, 2022), covering 46.2%, 55.9%, 42.2%, and 47.3% of the accommodations reported by the Japan Tourism Agency [13]. Each reservation record contains the reserved date, accommodation date, and cancelled date if the reservation is cancelled, and the reservation category listed in Table 1. Because the reservation records for the stay more than one year ahead is sufficiently rare (less than 0.0015% of all records), only the reservation records for the stay within 365 days are used for the analyses.

2.2 Reservation timing and cancellation probabilities

First, we focus on how many days the customers make a reservation in advance of the stay. The distribution of the new reservations x days prior to the stay features a monotonically reduction curve for x (see dots in Fig. 2). Based on this characteristic, we simply assumed the probability distribution of the reservations made x days ahead of the stay that could be approximated by the sum of exponential functions for x and maybe a constant term. Accordingly, it could be represented as

$$R_{K,O}(x) = \sum_{i=1}^{K} a_i \exp\left(-b_i x\right), \tag{1a}$$

$$R_{K,I}(x) = \sum_{i=1}^{K} a_i \exp\left(-b_i x\right) + c, \tag{1b}$$

where the model characteristics are represented by two subscripts, that is, the first subscript K denotes the number of exponential functions, and the second subscript (O or I) denotes the existence/absence of the constant term c.

We have assumed that the observed new reservation number during the focal time period follows the Poisson distribution whose expected occurrence number is the mean number of new reservations $\bar{r} = \sum_{t=1\ldots T} \sum_x R_{Data}(t, x)/T$, where T is the number of days from December 28, 2015 to December 29, 2019, and $x = 0\ldots365$ in our analyses. Accordingly, the likelihood function of $R_{K,Z}$ is as follows, where $Z \in \{O, I\}$:

$$L_R = \prod_t \prod_x pmf\left(\text{Poisson}\left(\bar{r} R_{K,Z}(x)\right), R_{Data}(t, x)\right), \tag{2}$$

where $R_{Data}(t, x)$ is the observed number of accommodation reservations for the stay x days ahead during the time period t, and $pmf(\text{Poisson}(k), m)$ denotes the probability mass functions of the Poisson distributions when the expected number of the event occurrence is k and the number of observed events is m. Note that $R_{K,Z}(x)$ determines the probability distribution of the new reservation for x days ahead and therefore $\sum_{x=0\ldots365} R_{K,Z}(x)$ is expected to equal one. For avoiding the complexity

Table 1 Reservation category contained in each reservation record

Name (Label)	Description	Classes
Guest type (GTY)	Composition of guests according to the number and gender of guests	Single: only one adult Couple: pair of adult male and adult female Family: containing child guest Others: all other cases
Guest number (GNM)	Number of guests (including both adults and children)	Single: only one guest Pair: two guests More: more than two guests
Guest number per room (GPR)	Whether the reserved room number is equal to the guest number	TRUE: equal FALSE: not equal
Stay length (LNG)	Length of the stay	One: stay one night Two: stay two nights More: stay more than two nights
Age class (AGE)	Age class of the reservation holder	Young: age < 30 Medium: 30 ≤ age < 50 Old: 50 ≤ age
With child (CHL)	Whether the guest younger than junior high school students (less than 13 years old) is included	With: at least a child is included in the guests Without: only adults in the guests
Accommodation room number (ARN)	Total room number in the accommodation	Small: #room < 100 Medium: 100 ≤ #room < 300 Large: #room ≤ 300
Accommodation distance (ADS)	Distance between the registered home address of the reservation holder and the accommodation address	Same: the accommodation located in the same prefecture as that of the consumer's home Close: the accommodation located in the prefecture sharing boundary with that of the consumer's home Far: all other cases

of the estimation with parameter restriction, in the following estimation we did not explicitly restrict $\sum_{x=0\ldots365} R_{K,Z}(x)$ as one, however, we confirmed that it is very close to one (0.9992 for the best model).

The cancellation probabilities exhibit two peaks, that is, one is immediately after the reservation made and the other is few days prior to the stay (dots shown in Fig. 5). Therefore, we assumed that the cancellation probability of the reservation that is reserved y days ahead of the stay and cancelled x days ahead of the stay, could be approximated by the sum of exponential functions for x and exponential functions for $y - x$ (that is, the number of days after the reservations). Accordingly, the cancellation probability of the existing reservations per day is represented as

$$C_{U,V,O}(x,y) = \sum_{i=1}^{U} d_i \exp\left(-e_i x\right) + \sum_{j=1}^{V} f_j \exp\left(-g_j\left[y - x\right]\right) \tag{3a}$$

$$C_{U,V,I}(x,y) = \sum_{i=1}^{U} d_i \exp\left(-e_i x\right) + \sum_{j=1}^{V} f_j \exp\left(-g_j\left[y - x\right]\right) + h \tag{3b}$$

where U denotes the number of exponential function of x, V denotes the number of exponential function of $y - x$, and the third subscript (I or O) denotes the existence/absence of the constant term h.

We assumed that the observed new cancellation numbers during the focal time periods follow the binomial distribution whose expected occurrence probability is $C_{U,V,Z}(x,y)$, where $Z \in \{O, I\}$ [Eqs. (3a) or (3b)]; the trial number is the number of 'surviving' (not cancelled yet) reservation at the beginning of the day. Based on these assumptions, the likelihood function of the model $C_{U,V,Z}$ is given by

$$L_C = \prod_t \prod_x \prod_y pmf\left(\text{Binomial}\left(N_{Data}(t,x,y), C_{U,V,Z}(x,y)\right), M_{Data}(t,x,y,)\right), \tag{4}$$

where $N_{Data}(t,x,y)$ is the surviving reservations x days ahead of the stay at the beginning of the day during the focal time period t that is the reservation y days ahead of the stay; $M_{Data}(t,x,y)$ is the cancelled reservations x days ahead of the stay during the time period t that is the reservation y days ahead of the stay, respectively. Because we used only the reservation records for the stay within 365 days, in our analyses, $y = 0\ldots365$ and $x = 0\ldots y$. $pmf(\text{Binomial}(n,p),m)$ denotes the probability mass functions of the binomial distributions when the number of observed events is m, the trial number is n, and the probability that an event occurs is p.

2.3 Estimation, cross-validation, and confidence interval

For the estimation, the observed number of accommodation reservations R_{Data}, the number of surviving reservations N_{Data}, and the number of the cancelled reservations M_{Data} are aggregated from all the given reservation records. The coefficients of the model for the reservation timing, $a_{1\ldots K}, b_{1\ldots K}$, and c in Eqs. (1a) and (1b), maximizing the likelihood L_R under the given subscripts K, Z, and R_{Data}

were estimated by the Nelder–Mead method. Similarly, the coefficients of the model for the cancellation probability per day, $d_{1\ldots U}, e_{1\ldots U}, f_{1\ldots V}, g_{1\ldots V}$, and h in Eqs. (3a) and (3b), maximizing the likelihood L_C under the given subscripts U, V, Z, M_{Data}, and N_{Data} were computed via the Nelder–Mead method.

For the comparison of the performance among models with different subscripts $K, U, V,$ or Z, we applied k-fold cross-validation, as described here. The original reservation records are partitioned by the week defined in ISO 8601 (that is, partitioned into 209 sub datasets). We calculated the coefficients maximizing the likelihood L_R and L_C estimated using all sub datasets other than the i-th sub datasets as well as the estimated coefficients; accordingly, the likelihood of the i-th sub datasets and its Bayesian information criterion (BIC) [24] are calculated. The performance of each model is evaluated by the mean value of the BIC of the i-th sub datasets for $i = 1 \ldots 209$.

The confidence interval of the model coefficients and the model estimation are computed by the bootstrap method as described here. The original reservation records are partitioned again into 209 sub datasets by week. The bootstrap sample is obtained by random sampling with replacement of the 209 sub datasets, and the model coefficients and the estimation of the model [that is, $R_{K,Z}(x)$ and $C_{U,V,Z}(x,y)$] are obtained by maximum likelihood estimation. By repeating this process 100,000 times, the confidence interval of the model coefficients and the estimated reservation timing/cancellation probabilities of the model are computed.

2.4 Determination of remarkable reservation category

We determined the remarkable reservation categories from that provided in the dataset shown in Table 1 for predicting reservation and cancellation by comparing the model performances with the different combinations of reservation categories. To assess model performances using goodness-of-fit statistics, for example BIC, the datasets between models using different combinations of reservation categories should be the same. Here, we explain preparation of the dataset for assessment. Consider the usage of "Guest Number" (GNM) category that classifies the number of guests in the focal record into three classes, that is, "Single," "Pair," or "More" than two guests. According to the GNM category, the reservation records are partitioned into three sub datasets, and for each sub dataset, the model parameters of the reservation timing can be estimated according to the maximum likelihood estimation explained above. Represent the estimates of the probability of reservation made from the sub dataset of class $\omega \in \{Single, Pair, More\}$ as $R_\omega^*(x)$, and that from the entire dataset (that is, the model estimate in the previous section) as $R^*(x)$. Note that the number of exponential functions K and the existence/absence of the constant term Z of $R_\omega^*(x)$ and $R^*(x)$ are chosen following the best model of the analyses described in the previous section.

Using $R_\omega^*(x)$, the new model using the GNM category, $\mathcal{R}'_{GNM}$, could be represented as

$$\mathcal{R}'_{GNM}(w,x) = \sum_{\omega \in \{Single, Pair, More\}} \delta_{\omega,w} R^*_\omega(x), \tag{5}$$

where $w \in \{Single, Pair, More\}$ is the GNM category of the focal reservation record, and $\delta_{i,j}$ is Kronecker delta:

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

Furthermore, the model that ignores the GNM category is

$$\mathcal{R}'_{nocategory}(w,x) = R^*(x). \tag{7}$$

The likelihood function of a model $\mathcal{R}'(w,x) \in \left\{\mathcal{R}'_{nocategory}, \mathcal{R}'_{GNM}\right\}$ for the reservation records partitioned by GNM category is given by

$$L_{\mathcal{R}'} = \prod_{w \in \{Single, Pair, More\}} \prod_t \prod_x pmf\left(\text{Poisson}\left(\bar{r}'(w)\mathcal{R}'(w,x)\right), \mathcal{R}'_{Data}(w,t,x)\right), \tag{8}$$

where $\mathcal{R}'_{Data}(w,t,x)$ is the number of new reservation for the stay x days ahead with GNM category w during the focal time period t, and $\bar{r}'(w)$ is the mean number of new reservations per class in GNM category w, that is, $\bar{r}'(w) = \sum_{t=1...T} \sum_x \mathcal{R}'_{Data}(w,t,x)/T$.

By applying the k-fold cross-validation, we can compare the model performance between $\mathcal{R}'_{no\ category}$ and $\mathcal{R}'_{GNM}$. To measure the model performance, BIC is calculated using the likelihood function $L_{\mathcal{R}}$. If the mean value of the BIC of $\mathcal{R}\prime_{GNM}$ in the k-fold cross-validation is smaller than that of $\mathcal{R}'_{GNM}$, the use of GNM category improves the model performance.

This method could be extended for the use of multiple reservation categories. We have eight reservation categories (Table 1); therefore, there are $2^8 = 256$ models depending on whether each of eight reservation categories are used for the estimation. We compare the performance of these 256 models as follows: represent the set of the classes in these reservation categories as $W_1, \dots, W_8$. Consider the model using $q_1 \dots q_N$-th of eight reservation categories. We represent the estimates of the probability of reservation made from the sub dataset of class $\omega_{q_1} \in W_{q_1}, \cdots, \omega_{q_N} \in W_{q_N}$ as $R^*_{\omega_{q_1}, \dots, \omega_{q_N}}(x)$. Accordingly, the model of the probability of reservation made x days ahead of the stay using the combination of $q_1 \dots q_N$-th reservation categories, $\mathcal{R}_{q_1 \dots q_N}$, is written as

$$\mathcal{R}_{q_1 \dots q_N}(w_1, \dots, w_8, x) = \sum_{\omega_{q_1} \in W_{q_1}} \cdots \sum_{\omega_{q_N} \in W_{q_N}} \left[\delta_{w_{q_1}, \omega_{q_1}} \times \cdots \times \delta_{w_{q_N}, \omega_{q_N}} \times R^*_{\omega_{q_1}, \dots, \omega_{q_N}}(x)\right], \tag{9}$$

where $w_1, \dots, w_8$ are the classes in each reservation category of the focal reservation record, that is, $w_1 \in W_1, \cdots, w_8 \in W_8$. The model that ignores any reservation category is

$$\mathcal{R}_{nocategory}(w_1, \dots, w_8, x) = R^*(x) \tag{10}$$

Accordingly, the likelihood function of a model for the probability of reservation made x days ahead of the stay with all reservation categories $\mathcal{R}(w_1, \ldots, w_8, x)$(which is either $\mathcal{R}_{nocategory}$ or $\mathcal{R}_{q_1 \ldots q_N}$) is expressed as

$$L_{\mathcal{R}} = \prod_{w_1 \in W_1} \cdots \prod_{w_8 \in W_8} \prod_t \prod_x pmf\left(Poisson\left(\bar{r}(w_1, \ldots, w_8)\mathcal{R}(w_1, \ldots, w_8, x)\right), R_{Data}(w_1, \ldots, w_8, t, x)\right),$$
(11)

where $\mathcal{R}_{Data}(w_1, \ldots, w_8, t, x)$ is the number of new reservations for the stay x days ahead with reservation category classes $w_1, \ldots w_8$ during the focal time period t, and $\bar{r}(w_1, \ldots, w_8)$ is the mean number of new reservations with reservation category classes $w_1, \ldots, w_8$, that is, $\bar{r}(w_1, \ldots, w_8) = \sum_{t=1 \ldots T} \sum_x \mathcal{R}_{Data}(w_1, \ldots, w_8, t, x)/T$.

Similarly, we represent the estimated cancellation probability of the existing reservations from the sub dataset of class $\omega_{q_1} \in W_{q_1}, \cdots, \omega_{q_N} \in W_{q_N}$ as $C^*_{\omega_{q_1}, \ldots, \omega_{q_N}}(x, y)$, and that from the entire dataset (that is, the model estimate in the previous section) as $C^*(x, y)$. Again, note that the number of exponential functions U, V and the existence/absence of the constant term Z of $C^*_{\omega_{q_1}, \ldots, \omega_{q_N}}$ are chosen following the best model of the analyses described in the previous section. Accordingly, the model of the probability of cancellation x days ahead of the stay cancelling y days of the stay, using $q_1 \ldots q_N$-th reservation categories, $C_{q_1 \ldots q_N}$, is written as

$$C_{q_1 \ldots q_N}(w_1, \ldots, w_8, x, y) = \sum_{\omega_{q_1} \in W_{q_1}} \cdots \sum_{\omega_{q_N} \in W_{q_N}} \left[\delta_{w_{q_1}, \omega_{q_1}} \times \cdots \times \delta_{w_{q_N}, \omega_{q_N}} \times C^*_{\omega_{q_1}, \ldots, \omega_{q_N}}(x, y)\right],$$
(12)

where $w_1, \ldots w_8$ are the reservation category classes of the focal reservation record, that is, $w_1 \in W_1, \cdots, w_8 \in W_8$. The model that ignores any reservation category is

$$C_{nocategory}(w_1, \ldots, w_8, x, y) = C^*(x, y) \tag{13}$$

Accordingly, the likelihood function of a model for the probability of reservation made x days ahead of the stay with all reservation categories $C(w_1, \ldots, w_8, x, y)$ (which is either $C_{nocategory}$ or $C_{q_1 \ldots q_N}$) is expressed as

$$L_C = \prod_{w_1 \in W_1} \cdots \prod_{w_8 \in W_8} \prod_t \prod_x$$
$$pmf\left(Binomial\left(N_{Data}(w_1, \ldots, w_8, t, x, y), C(w_1, \ldots, w_8, x, y)\right), M_{Data}(w_1, \ldots, w_8, t, x, y)\right),$$
(14)

where $N_{Data}(w_1, \ldots, w_8, t, x, y)$ is the surviving reservations whose reservation category classes are $w_1, \ldots, w_8$ x days ahead of the stay at the beginning of the day during the focal time period t that is the reservation y days ahead of the stay, and $M_{Data}(w_1, \ldots, w_8, t, x, y)$ is the cancelled reservations whose reservation category classes are $w_1, \ldots, w_8$ x days ahead of the stay during the time period t that is the reservation y days ahead of the stay.

Applying the k-fold cross-validation, the mean value of the BIC of $\mathcal{R}$ and C derived from the likelihood function $L_{\mathcal{R}}$ and L_C, respectively, in the k-fold cross-validation were calculated and compared among 256 models, that is, the possible combinations of reservation categories. We determined the remarkable factors as the

reservation category in the model whose mean BIC in the k-fold cross-validation is minimum.

3 Result

3.1 Estimation of the model for the reservation timing

First, we show the results of the estimation of the model for the reservation timing. Figure 1 shows the mean BIC of the model $R_{K,Z}$ with the variation in the number of exponential function K (horizontal axis) and the existence/absence of a constant term (red and blue, respectively). Regardless of the presence of the constant term, the mean BIC exhibits a clear valley shape whose bottom is $K=5$, where the model without constant term features a smaller mean BIC than that with a constant term. Note that the mean negative log-likelihood shows a gradually decreasing trend, but is mostly flat at $K>5$ (Figure S1), indicating the sufficiently small accuracy improvement of the model estimation upon increasing the number of exponential functions at $K \geq 5$.

The estimation of the best model $R_{5,O}$, that is, the five exponential functions without any constant term, (red curve in Fig. 2) show suitable fitting with the observed data (black dots on Fig. 2). The shape of the reservation timing seems to be divided into three phases, that is, $x<60$, $60 \leq x < 180$, and $180 \leq x$. For the reservation made more than a half year ahead of the stay tend to exponentially increase with the days until the stay (Fig. 2b). The exponential growth of the reservation made is the same more than two months but less than a half year ahead of the stay, although the growth rate slows down (the slope becomes gentle at $x<180$, as shown in Fig. 2b). Furthermore, the new reservation within two months until the stay exhibits faster growth than the other periods, and this growth continues until $x=0$, that is, the probability of reservation made is highest for the stay on the same day.
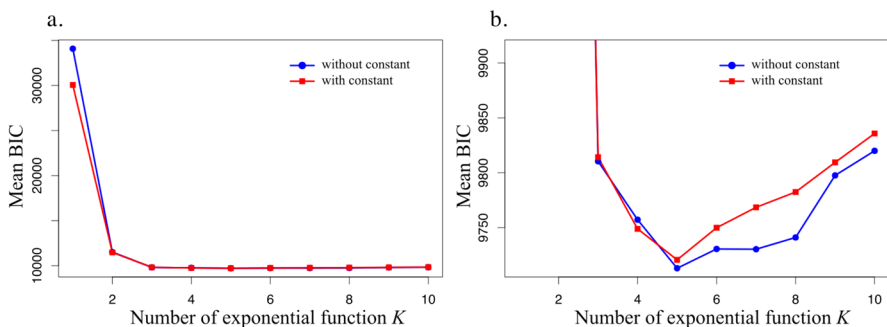

Fig. 1 Performance of the reservation timing models. The mean BIC in the k-fold cross-validation for the model of the probability of reservations made x days ahead of the stay (panel **a**), and that with enlarging the y-axis (panel **b**). Horizontal axis shows the number of exponential functions K, and the colours indicate the existence/absence of a constant term $Z \in \{I, O\}$ (red and blue, respectively)
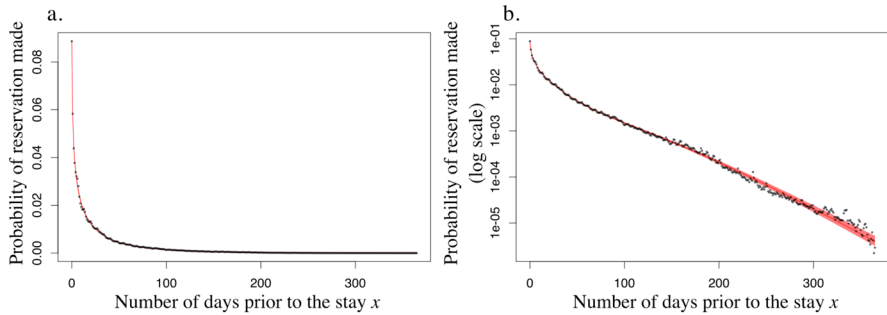
Fig. 2 Probability of the days of the reservations until the stay. **a** the data (black dots) and the estimation by the best model $R_{5,O}$ (red curve) of the probability of reservations made x days ahead of the stay, and **b** that with log-scale vertical axis. Pale red band indicates the 99% confidential interval computed by the bootstrap method

According to the estimated coefficients of the best model (Table S1), the 1st, 2nd, and 4th exponential functions exhibit similar exponents ($0.0283 \leq b_i \leq 0.0364$), while the exponents of the other two exponential functions are different in terms of the magnitude of order. Because the sum of exponential functions with similar exponents with each other also shows a similar trend, hereafter we classify five exponential functions into three groups, each of which is called as 'Short-K' (5th function, $b_5 = 1.389$, so the mean days until the stay is 0.78), 'Middle-K' (3rd function, $b_3 = 0.2835$, so the mean days us 3.53), or 'Long-K' (1st, 2nd, and 4th function, $\{b_1, b_2, b_4\} = \{0.028, 0.032, 0.036\}$, so the mean days are 35.2, 31.2 and 27.5), respectively (see Table S1). This classification can also be supported by the comparison of the model with less than five exponential functions (Figure S2), that is, the increment of the number of the exponential functions drastically improves the fitting until $K \leq 3$, while the improvement of fitting seems to be limited for $x < 180$ at $K = 4$ and 5, suggesting that the general trend tends to be fitted by three exponential functions.

Figure 3 shows the values of three groups of exponential functions. The growth of Short-K becomes remarkable on the reservation for the stay within next 3 days, while the growth of Middle-K becomes remarkable on the reservation for the stay within the next 3 weeks. The contribution of the Short-K, Middle-K, and Long-K account for one-third of the reservation for the stay on the same day individually. Interestingly, the estimated values of each of the three groups are highly robust (the 99% confidential interval is sufficiently small in Fig. 3b).

3.2 Estimation for the model of cancellation probability

Next, we show the results of the estimation of the model for the cancellation probability $C_{U,V,Z}$. Figure 4 shows the mean BIC of the model $C_{U,V,Z}$ with the change in the number of exponential functions of the days until the stay V (horizontal axis), the number of exponential functions of the days from the reservation U (vertical axis), and the existence/absence of the constant term (panel a and panel b). The mean BIC in the k-fold cross-validation becomes minimum at $C_{4,2,O}$, that is, four exponential
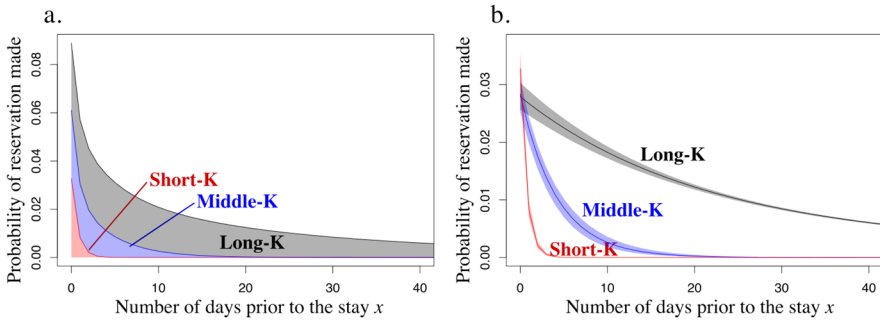
Fig. 3 Exponential functions composing the best model of reservation timing. The stacked line plot (panel **a**) and the line plot (panel **b**) of the three groups of exponential functions. In panel **b**, the pale bands indicate the 99% confidential interval computed by the bootstrap method

functions of the days until the stay, two exponential functions of the days from the reservation without constant term. Note that the mean negative log-likelihood shows an almost flat shape when $U > 4$ and $V > 2$ (Figure S3), indicating that the accuracy improvement of the model estimation upon increasing the number of exponential functions becomes sufficiently small at $U > 4$ or $V > 2$.

Unlike the model of the reservation timing, the models of cancellation probability depend on two variables—the days until the stay x and the days from the reservation made to the stay y. Figure 5 shows the estimation of the best model $C_{4,2,O}$ (curves) and the observed data (dots) for different y values (colors). The estimated results clearly show two-peaks pattern of the cancellation probability, that is, immediately after the reservation made (that is, $y = x$), the cancellation probabilities become high, but it declines rapidly within the next three weeks. This pattern is common, regardless of the days until the stay from the reservation y. Furthermore, the cancellation probability also increases with the days until the stay x decreases. This cancellation growth approaching the stay reaches
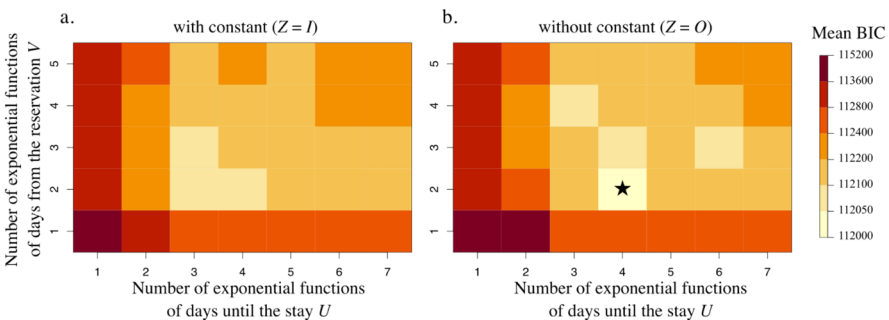
Fig. 4 Performance of the cancellation probability models. The mean BIC in the k-fold cross-validation for the model of the probability of cancellation x days ahead of the stay that is reserved y days ahead of the stay. Star mark indicates the best model, that is, the model whose mean BIC is the minimum
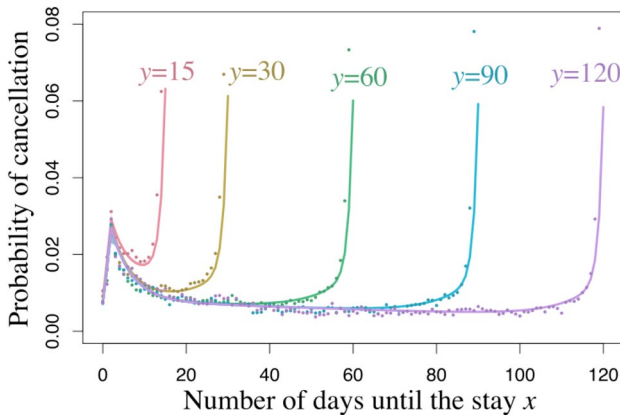
Fig. 5 Daily cancellation probability of the reservation x days ahead of the stay reserving y days of the stay. The dots show the data of the cancellation probability of the reservation that is reserved y days ahead of the stay and cancelled x days ahead of the stay, and lines shows the predictions of the best model $C_{4,2,O}$. Note that the cancellation probability is two-dimensional (i.e., x and y), we plotted only for selected five y values (y = 15, 30, 60, 90 and 120) indicated by colors of dots and lines. Pale band indicates the 99% confidential interval computed by the bootstrap method

a peak around few days before the stay, and declines rapidly almost immediately before the stay.

According to the estimated coefficients of the best model (Table S2), the 1st and 4th exponential functions of the days until the stay have similar exponent ($e_1 = 1.109$ and $e_4 = 1.107$), while the exponents of the other exponential functions have a different magnitude of order. Therefore, hereafter we classify four exponential functions of the days until the stay into three groups, each of which is called "Short-U" (the 1st and 4th functions), "Middle-U" (the 3rd function), or "Long-U" (the 2nd function), respectively (see Table S1). Similarly, we call the first and second exponential functions of the days from the reservation made as "Short-V" and "Long-V," respectively.

Figure 6a shows the values of the five groups of exponential functions. The cancellation immediately after the reservation could be represented by two exponential functions; one exhibits remarkable influence until the next 5 days after reservation, while the other maintains the remarkable influence for approximately three weeks. Furthermore, the cancellation nearing the stay is represented by three groups; Long-U shows an extremely slow reduction of the cancellation rate, while Middle-U shows a remarkable influence on the cancellation rate within 2 weeks. Interestingly, the coefficient of Short-U is negative, that is, this group contributes to the decrease in the cancellation rate within few days until the stay, causing the peak of the cancellation probability to approach the stay few days before the stay. The estimated values of each of the five groups are highly robust (the 99% confidential interval is sufficiently small, as shown in Fig. 6b and c).
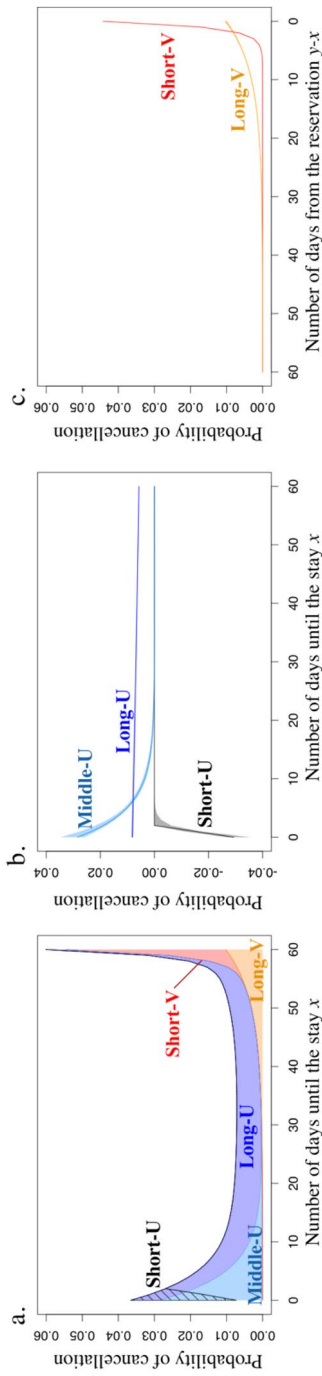
Fig. 6 Exponential functions composing the best model of cancellation probability. **a** The stacked line plot (panel **a**) of five groups of exponential functions, when $y = 60$. Value of the Short-U (hatched region) is negative, indicating the realized cancellation probability of the values removing the hatched region from the stack of the other groups. **b** The values of Short-U, Middle-U, and Long-U x days ahead of the stay. **c** The values of Short-V and Long-V y days since the reservation is made. Pale bands in panel b and c indicate the 99% confidential interval computed by the bootstrap method

3.3 Determination of the remarkable reservation category

Figures 7 and 8 show the performance of the model in terms of the reservation timing and cancellation probability with different combinations of reservation categories. For the reservation timing, the model using three reservation categories, that is, Guest type (GTY), Guest number per room (GPR), and Stay length (LNG), shows the minimum mean BIC in the k-fold cross-validation. For the cancellation probability, the model using only GPR shows the minimum mean BIC in the k-fold cross-validation. In both analyses, the mean negative log-likelihood always decreases upon adding a new category to the model, indicating that the accuracy of the model estimation by adding reservation category always improves, although the improvement is not sufficient for the penalty based on the increment of the model parameters (Figure S4 and S5).
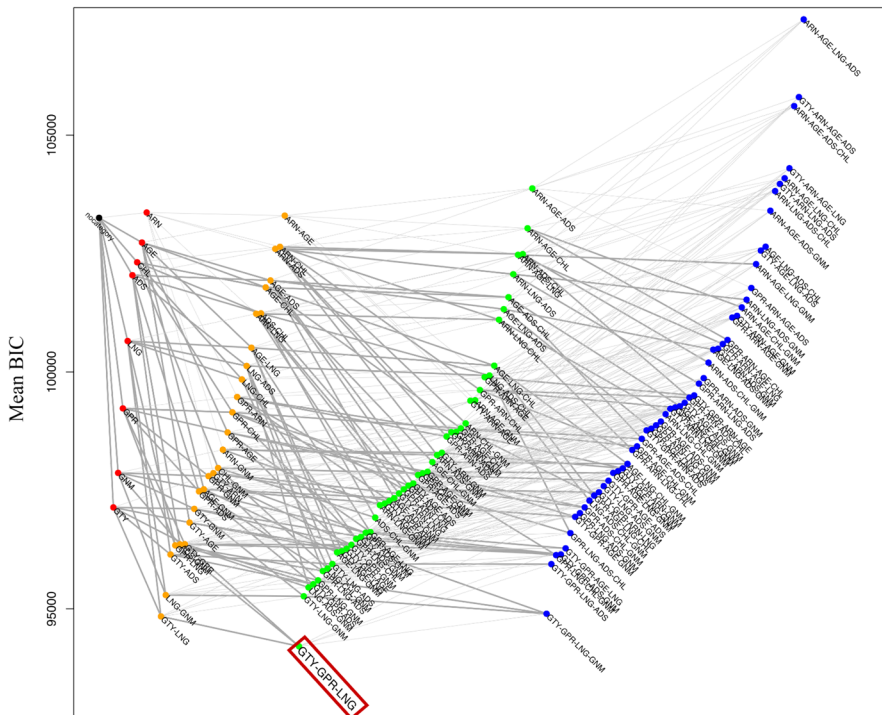
Fig. 7 Model performance using different combinations of reservation category for the model of reservation timing. Each points indicate the mean BIC of the models using different combinations of reservation categories labelled next to the points. Colors indicate the number of different categories, and the lines indicate that the right model can be generated by adding a new category to the left model. The thickness of the lines indicates whether such an additional category can improve the model performance; thicker line improve
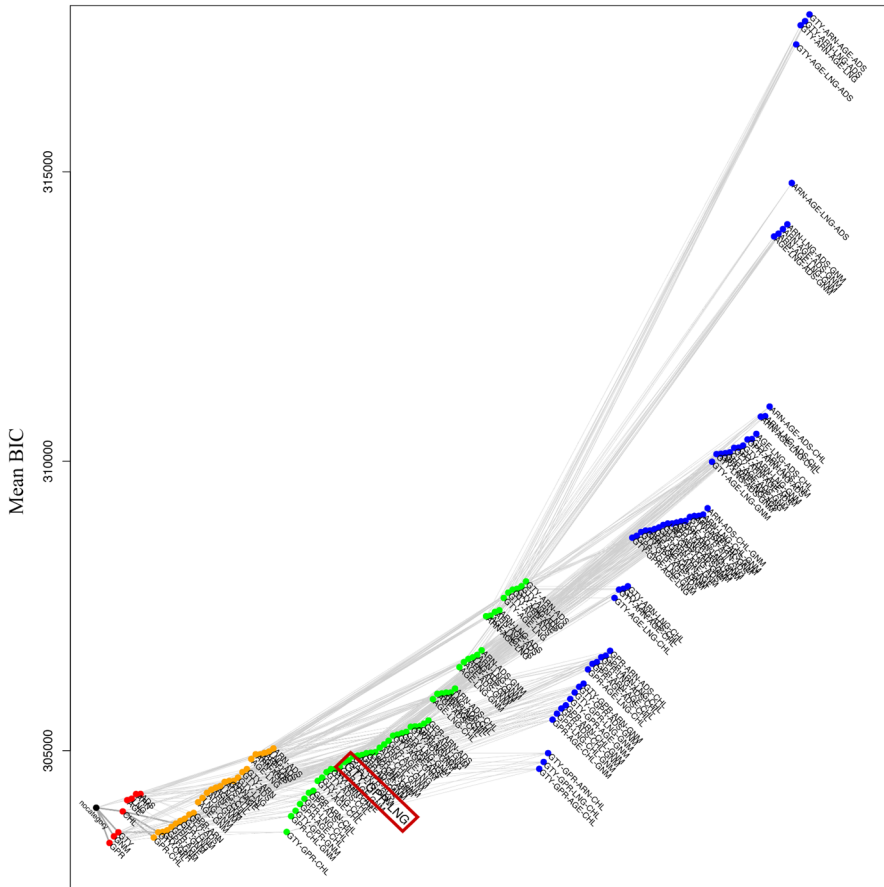
Fig. 8 Model performance using different combinations of reservation category for the model of cancellation probability. The meaning of lines and colors are the same as that in Fig. 7. Note that the models using more than four reservation categories are not shown, although their mean BIC is higher than that of the best model

4 Discussion

In this study, we revealed the general trends in the time changes of the customer behavior on the accommodation reservations. Regardless of the importance of the basic knowledge of customer behaviors for forecasting the revenue management, the general behavioral patterns of the customers on accommodation reservation have not been adequately studied. We showed that in the best model, the reservation timing was decomposed into five exponential functions of the days until the stay, and the cancellation probability was decomposed into four exponential functions of the days until the stay and two exponential functions of the days from the reservation. We also chose the guest type, guest number per room, and stay length as the reservation categories that exhibit remarkable influence on the reservation

timing model, while only the guest number per room was chosen for the cancellation probability model.

Our results afford important suggestions regarding the customer behavior during accommodation reservation. In the estimation of the reservation timing, the exponents of the exponential functions were categorized into three groups (short-K, middle-K, and long-K), suggesting that there are three different trends of making reservations. These three trends could be viewed as a reflection of the difference in the importance of the trips. For example, highly important schedules such as appointments with important business partners or family vacation are schedules defined in advance at the earliest possible. Furthermore, less important trips are scheduled after those whose schedules do not conflict with the more important schedules, that is, when there is a vacancy in the schedule. Hence, the reservations for the less important schedules such as the solo short leisure travels or non-urgent business visits tend to be scheduled as the schedule date approaches. The three trends of the reservation timing might reflect that people classify their schedules into three levels according to the importance. Accordingly, the exponents of exponential functions suggest that decision making for the moderately important schedules is completed within three weeks ahead of the date, while that for the less important schedules are not made until three days before the stay.

Another possible explanation of the three trends in the reservation timing is the availability of alternative choices on the reservation. For example, if the trip purpose is to stay in an accommodation with unique features, or if the trip is scheduled during the famous festival in the town, or if there are few accommodations in the focal town, it is difficult to find alternative accommodations. Therefore, the reservation might be made far ahead of the stay. Furthermore, in big cities we can easily find alternative hotels, particularly for business purpose, and therefore, the customers might tend to make their reservations immediately before the stay. The three trends might reflect the categorization of such availability into three levels.

The remarkable reservation categories in the reservation timing model (Fig. 7) might support the explanations presented above. The best model using the reservation categories was the guest type, guest per room, and stay length. The guest type tends to be related to the trip purpose, that is, leisure purpose must be common for couple or family classes, while the business purpose seems to be most appropriate for single or other classes. Guest per room is also related to the business/leisure purpose, that is, one room for each person is common in trips within the business context. Such difference of trip purpose changes the importance of the trip and/or the availability of alternative choices. Similarly, the stay length seems to correlate with the importance of the trip and the substitution possibility of accommodations because a long-stay trip increases the probability of the accommodation being full on any day during the trip.

Similar discussions could be extended to explain the customers' cancel behavior. First, the cancellation probability is explained appropriately by the sum of two types of exponential functions, that is, the functions of the days from the reservations and those of the days until the stay. This suggests that (i) the factors causing cancellation would be different between the cancellations immediately after the reservations and those immediately before the stay, and (ii) the

cancellations can be explained only by these two factors regardless of the days between the reservation made and the stay. In other words, the customers' cancellation behaviors are independent of the number of days ahead of the stay that the reservation made.

The best model of the cancellation probabilities contains two exponential functions of the days from the reservations, Short-V and Long-V, suggesting that the cancellations immediately after the reservations exhibit two different trends. One possible explanation of the cancellation could be attributed to whether the customers themselves book the reservation. Once a reservation is made, the customers may cancel them for various reasons, such as, mistakes in the reserved plan or date, change to another better accommodation found after the reservation, cancellation of vacations owing to business reasons, or failure of schedule adjustment of the traveling companions. Importantly, the cancellations due to the first two reasons are caused whilst the customers focus on planning the trip, while the cancellations owing to the last two reasons are caused after the customers share the travel information with others. The difference of exponents between Short-V and Long-V can be explained by the fact that the cancellation caused by the customers themselves tends to occur within a short time (mostly within the next five days in Short-V) because the customers focus on the planning of the trip only for such durations, while the cancellations caused by others than customers exhibit a longer trend because the information of the trip shared after planning is mostly finished.

Furthermore, the best model of the cancellation probabilities contains three groups of exponential functions of the days until the stay, Short-U and Middle-U, and Long-U, suggesting that the cancellations immediately before the stay exhibits three different trends. Remarkably, Short-U shows a negative value regardless of the days until the stay, that is, Short-U represents the influence of reduction of the cancellation probability on the days within few days of the stay. This seems to be explained by the cancellation charge. Although the detail of the change amount is different among the accommodations, most accommodations impose a cancellation charge that increases particularly immediately before the stay. Short-U seems to represent the reduction of the cancellation for avoiding such a high cancellation charge.

The other two groups, Middle-U and Long-U, represent the increment of the cancellation probability as the stay date approaches, although the growth rate of Long-U is extremely slow (twice for approximately 92 days). These two trends with different growth rates could be explained by the difference of the cancellation reasons. First, the accommodation reservation might be cancelled due to unexpected events that disturb the travel independent of the intension of customers, for example, natural or human disasters, injury or illness, or conflict with other more important schedules. Second, the accommodation reservation might also be cancelled owing to rescheduling by customers. For example, the customers might reconsider the schedule of trip in response to weather forecast, health condition, or reassessment of the travel requirements. Because unexpected events could occur regardless of the days until the stay, while the rescheduling of trip generally occurs after the expectation of forecasts, the health conditions or the other tasks of customers are clear. Such difference might be reflected in the exponential functions with different components, that is, Long-U and Middle-U.

Moreover, the exponent of Middle-U ($e_3 = 0.2962$) is interestingly similar to that of Middle-K ($b_3 = 0.2385$). This might represent the general pattern of the people managing their own schedules. According to our hypotheses, the customers tend to reconsider their own schedules over the next few weeks that increase the cancellation probability within the next three weeks from the stay represented as Middle-U. Simultaneously, such reconsideration of the schedule causes the planning of other new trips that increase the booking of reservations within the next three weeks represented as Middle-K. In other words, three weeks might be the limit of the duration within which people can accurately predict and consider their own future.

Our analyses of the reservation categories reveal the important factors influencing the customer behavior on accommodation reservation. In the models for both reservation timing and cancellation probability, GPR has been selected in the best model including the reservation categories. As we discussed above, GPR can be considered as the indicator of the relationship among guests, that is, the sharing of one room with multiple guests suggests friends or families, while the use of one room per person suggests stay with less-related persons, particularly in business. This suggests that one of the most critical categories influencing the customers' behaviors on both booking and cancelling reservation is whether the travel purpose is leisure or business. Although the guest type (GTY) was selected in the best model for reservation timing, but not selected in that for cancellation, the single inclusion of the guest type exhibits suitable improvement of the model BIC, suggesting that the cancellation behaviors can be explained by the guest type suitably. Contrary to these, the stay length (LNG) is selected in the best model for reservation, but in the cancellation model, the inclusion of the LNG category type degrades the mean BIC. When the vacancy of the accommodations reduces, the reservation of long stay becomes considerably more difficult than that for short stay because of the determination of the accommodation that has vacancy for all days of the stay. This might be the reason why the proportion of the timing of booking the reservation is strongly influenced by the stay length. Furthermore, the result that the LNG category is not selected indicates that the cancellation reason is independent of the stay length.

Moreover, the solitary inclusion of accommodation distance (ADS) or accommodation room number (ARN) improves negligibly or deteriorates the mean BIC of both reservation models and cancellation models. These suggest that the accommodation features have negligible influence on the decision making of customers. Similarly, the age of the reservation holders (AGE) and the existence/absence of a child (CHL) also exhibit negligible improvement of the BIC of the models. Interestingly, these factors are strongly related with the guest type (GTY) that is selected as one of the three reservation categories for reservation model. Particularly, stay with the child is the definition of one of the GTY class "family." This implies that the distinction of whether the focal reservation is made by family is not sufficient for improving the model performance, that is, there is a sufficiently large difference in behaviors during booking reservations among a single person, a couple, and other cases involving families.

The best model of the reservation probability contains five exponential functions, three of which are categorized into Long-K. The role of these three functions is the improvement of the fitting to the waving shape of the observed data

in $x > 150$, for example, the observed data seems to be higher around $x = 180$ and 320 and lower around $x = 240$ compared with the exponential growth (Fig. 2b). These waving shapes might reflect the interval of the planning for the next travels, that is, a relatively higher proportion of reservation around $x = 180$ and 320 might be the reservation for the planning of the summer/winter vacation on winter/summer or the vacation planning after the end of the vacation on the same season.

Although our analyses revealed the mass-level behaviors of customers during accommodation reservation; this is not sufficient for fully understanding the individual-level behaviors. For example, customers might often cancel an accommodation and make a reservation for another accommodation on the same day; but in our analysis, these two behaviors are recorded as two independent events, that is, one cancellation and one reservation made. In future work, further analyses of the reservation records linking with the user ID on PNR will be required for revealing the general trends of individual-level behaviors on the accommodation reservation.

Another important future work is the comparison of the customer behavior on the reservation focusing on different services. For example, the application of our method into the reservation records of other than hotel industries, for example, flight or rail tickets, healthcare, or restaurants, might exhibit different time change patterns on the customer behavior. The comparison of the customer behavior among different industries reveal human decision-making patterns pertaining to scheduling, and the response of customers to the difficulty of alternative choices, penalty of cancellation, or the relative importance of those appointments. Indeed, Romero Morales and Wang [22] reported the time change of the cancellation probability depending on the days until the service provided in the airline ticket, but there results exhibit a gradual decreasing shape as the flight date approaches. This might represent the importance or the difficulty of alternative choices of flight tickets compared with the accommodation reservations. Another possible future work will be the influence of the culture or nationality on the reservation behaviors. Although the OTA providing dataset for this study mainly provides its service in Japan, the comparison with the dataset of OTA on different countries might reveal the influence of culture or nationality on the decision-making patterns.

In conclusion, we revealed the general trends of the customer behaviors on the time change of the reservation made and cancellation. We have shown that the customer behavior on the accommodation reservation could be categorized into multiple factors motivating reservations or cancellations. These findings facilitate a deeper understanding of the decision-making patterns of customers that could contribute to constructing accurate forecasting models in the hotel industry. Moreover, our results contribute to the development of a general forecasting model that shall be widely applicable to various accommodations and situations.

Declarations

References

1. Adams, W., Vodicka, M., 1987. Short-term forecasting of passenger demand and some applications in Qantas. Presented at the Proceedings of the 27 Annual AGIFORS Symposium.
2. Aliperti, G., Sandholz, S., Hagenlocher, M., Rizzi, F., Frey, M., Garschagen, M.: Tourism, crisis, disaster: an interdisciplinary approach. Ann. Tour. Res. **79**, 102808 (2019). https://doi.org/10.1016/j.annals.2019.102808
3. Antonio, N., de Almeida, A., Nunes, L.: Predicting hotel booking cancellations to decrease uncertainty and increase revenue. TMS **13**, 25–39 (2017). https://doi.org/10.18089/tms.2017.13203
4. Antonio, N., de Almeida, A., Nunes, L.: Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. Cornell Hospit. Q. **60**, 298–319 (2019). https://doi.org/10.1177/1938965519851466
5. Beckmann, M.J., Bobkoski, F.: Airline demand: an analysis of some frequency distributions. Nav. Res. Logist. Q. **5**, 43–51 (1958). https://doi.org/10.1002/nav.3800050105
6. Carrier, E., Weatherford, L.: Estimating a model of airline passenger choice using grouped booking data and least squares regression. J. Revenue Pricing Manag. (2014). https://doi.org/10.1057/rpm.2014.19
7. Chow, W.S., Shyu, J.-C., Wang, K.-C.: Developing a forecast system for hotel occupancy rate using integrated ARIMA models. J. Int. Hosp. Leisure Tour. Manag. **1**, 55–80 (1998). https://doi.org/10.1300/J268v01n03_05
8. El Haddad, R., Roper, A., Jones, P., 2008. The impact of revenue management decisions on customers attitudes and behaviours: A case study of a leading UK budget hotel chain, in: EuroCHRIE 2008 Congress, Emirates Hotel School, Dubai, UAE, 11th-14th October. Retrieved April. p. 2011.
9. Falk, M., Vieru, M.: Modelling the cancellation behaviour of hotel guests. Int. J. Contemp. Hosp. Manag. **30**, 3100–3116 (2018). https://doi.org/10.1108/IJCHM-08-2017-0509
10. Hang, H., Aroean, L., Chen, Z.: Building emotional attachment during COVID-19. Ann. Tour. Res. **83**, 103006 (2020). https://doi.org/10.1016/j.annals.2020.103006

11. Ilgaz, B., Dundar, Y., Silik, C.E., Oral, M.A.: The impact of the covid-19 outbreak on local tourist behavior in Antalya, Turkey. Sci. Pap.-Ser Manag. Econ. Eng. Agric. Rural Dev. **21**, 391–403 (2021)

12. International Civil Aviation Organization: Guidelines on Passenger Name Record (PNR) Data (2010). https://www.icao.int/Security/FAL/ANNEX9/Documents/9944_cons_en.pdf

13. Japan Tourism Agency, 2021. Accommodation and Travel Statistics Survey (November 2021; only available in Japanese).

14. Japan Travel and Tourism Association, VALUES, Inc., 2022. Travel and tourism websites views ranking in 2020.

15. Lee, A.O., 1990. Airline reservations forecasting : probabilistic and statistical models of the booking process (Technical Report). Cambridge, Mass. : Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology, [1990].

16. L'heureux, E., 1986. A new twist in forecasting short-term passenger pickup. Presented at the Proceedings of the 26th Annual AGIFORS Symposium.

17. Martínez-Ibarra, E.: Climate, water and tourism: causes and effects of droughts associated with urban development and tourism in Benidorm (Spain). Int. J. Biometeorol. **59**, 487–501 (2015). https://doi.org/10.1007/s00484-014-0851-3

18. McGill, J.I.: Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints. Ann. Oper. Res. **60**, 209–240 (1995). https://doi.org/10.1007/BF02031947

19. McGill, J.I., Van Ryzin, G.J.: Revenue management: Research overview and prospects. Transp. Sci. **33**(2), 233–256 (1999). https://doi.org/10.1287/trsc.33.2.233

20. Pappas, N., Glyptou, K.: Accommodation decision-making during the COVID-19 pandemic: complexity insights from Greece. Int. J. Hosp. Manag. **93**, 102767 (2021). https://doi.org/10.1016/j.ijhm.2020.102767

21. Ritchie, B.W., Jiang, Y.: A review of research on tourism risk, crisis and disaster management: launching the annals of tourism research curated collection on tourism risk, crisis and disaster management. Ann. Tour. Res. **79**, 102812 (2019). https://doi.org/10.1016/j.annals.2019.102812

22. Romero Morales, D., Wang, J.: Forecasting cancellation rates for services booking revenue management using data mining. Eur. J. Oper. Res. **202**, 554–562 (2010). https://doi.org/10.1016/j.ejor.2009.06.006

23. Sánchez-Medina, A.J., C-Sánchez, E.: Using machine learning and big data for efficient forecasting of hotel booking cancellations. Int. J. Hosp. Manag. **89**, 102546 (2020). https://doi.org/10.1016/j.ijhm.2020.102546

24. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)

25. Smith, B.C., Leimkuhler, J.F., Darrow, R.M.: Yield management at American airlines. Interfaces **22**, 8–31 (1992). https://doi.org/10.1287/inte.22.1.8

26. Uglis, J., Jęczmyk, A., Zawadka, J., Wojcieszak-Zbierska, M.M., Pszczoła, M.: Impact of the COVID-19 pandemic on tourist plans: a case study from Poland. Curr. Tour. (2021). https://doi.org/10.1080/13683500.2021.1960803

27. Vulcano, G., van Ryzin, G., Chaar, W.: Choice-based revenue management: an empirical study of estimation and optimization. M&SOM **12**, 371–392 (2010). https://doi.org/10.1287/msom.1090.0275

28. Weatherford, L.: The history of forecasting models in revenue management. J. Revenue Pricing Manag. **15**, 212–221 (2016). https://doi.org/10.1057/rpm.2016.18

29. Weatherford, L.R., Kimes, S.E.: A comparison of forecasting methods for hotel revenue management. Int. J. Forecast. **19**, 401–415 (2003). https://doi.org/10.1016/S0169-2070(02)00011-0

30. Weatherford, L.R., Kimes, S.E., Scott, D.A.: Forecasting for hotel revenue management. Cornell Hotel Restaur. Adm. Q. (2001). https://doi.org/10.1177/0010880401424005

31. Zaki, H.: Forecasting for airline revenue management. J. Bus. Forecast. Methods Syst. **19**(1), 2–6 (2000)