International Journal for Research in Tourism and Hospitality
Published by: **Green Publication Services**
Volume 01 Issue 01 March 2026

DETERMINANTS OF HOTEL BOOKING CANCELLATIONS: EVIDENCE FROM RESERVATION ANALYTICS

Viraj Singh Rathod
Research Scholar, Jamia Millia Islamia, rathodvirajsingh@gmail.com

ABSTRACT

Hotel booking cancellations create substantial operational and financial challenges by disrupting demand forecasting, room inventory allocation, pricing strategies, and revenue management, while also affecting staff morale, service efficiency, and customer satisfaction. The present study explored the determinants of cancellation of hotel bookings by analysis of hotel reservations and a quantitative explanatory design. Characteristics of the customer, booking, pricing, distribution, and characteristics of the hotel were analysed. Descriptive statistics, bivariate tests and binary logistic regression were used to determine the significant predictors and model performance. The results indicated that customers' booking lead time, cancellation history, non-refundable deposit, average daily rate, group size and selected market segment were factors that positively influenced their likelihood of cancellation. On the other hand, prior non-cancelled bookings, resort hotel classification, resort modification and special requests were found to correlate with reduced risk of cancellation. The logistic regression model had a good predictive performance, with the accuracy achieved being more than 81% and the area under the receiver operating characteristic curve being 0.866. These results suggest that the information provided at the reservation level can be used effectively in order to separate high-risk bookings from more stable bookings. On a management level, hotels can take the cancellation risk estimates to enhance overbooking decisions, prioritise confirmation messages, make more accurate demand estimates and minimise revenue leakage. The study concludes that interpretable reservation analytics provides a viable framework to support evidence-based booking management.

Article History:

Article Type: **Review**

Received Date: **13/02/2026**
Revised Date: **14/03/2026**
Accepted Date: **21/03/2026**
Published Date: **29/03/2026**

Keywords: hotel booking cancellations; reservation analytics; logistic regression; revenue management; cancellation prediction

1. Introduction

There has been a growing concern regarding hotel booking cancellations as they affect the accuracy of demand forecasts, the ability to allocate rooms to bookings, and pricing, staffing and revenue determinations. Online booking has made bookings more convenient, but also offered greater flexibility, price comparison and booking cancellations for customers. Now travelers think about several options, are quick to react to the price changes, and adjust plans before they arrive. This behavior is a result of the relationship between the perceived value, booking conditions, booking timing and the characteristics of the channel chosen by the consumer when deciding on a hotel booking (Boto-García et al., 2021; Masiero et al., 2020). As such, the management of hotel cancellations is now a key part of hotel revenue management.

Cancellation policies are tailored to meet the needs of the hotel while providing convenience for the customer. Flexible policies can drive sales, but also may drive over-bookings, and restrictive policies can deter customers and/or be less competitive. Moderate cancellation policies have been shown to have better effects than overly liberal policies or overly restrictive policies (Altin et al., 2023). The monetary damage can also be felt beyond the loss of a reservation, as rooms that are cancelled could either be left unsold or sold at a reduced price or last-minute price changes may be costly. Another type of revenue leakage can happen when customer cancellations and rebooking happen after seeing lower fares, as this poses a problem for rate integrity and inventory control (Kmiecik & Antonio, 2026).

Various factors, such as behavioural, transactional, temporal & operational factors, affect booking cancellations. Leeway time may also add to the uncertainty because customers who make bookings have a longer period in which to change their travel plans. Online travel agencies can also produce unique cancellations, and clients can make several bookings at a time and easily compare hotels. Recent studies revealed that the timing and manner of cancellations differ according to online travel agency segments, suggesting that, in addition to investigating whether a booking is cancelled, it is important to consider the circumstances under which the cancellation is made (Atabay et al., 2026). Recent theory focuses on strategic choice, perceived risk, flexibility, and the evolving travel situation to explain cancellation behaviour (Mir et al., 2026).

The accurate cancellation analysis is thus essential for hotel demand forecasting. In cases where customer behaviour shifts quickly over time, across channels and customer segments, traditional forecasting methods might not be enough. The integration of machine learning and revenue management enables the hotel to continually fine-tune predictions, identify nonlinearities, and make better decisions when dealing with uncertainty (Das et al., 2023). Hotel demand forecasting has seen AI increasingly used, but the effectiveness of these models is linked to the quality of the data, model interpretability, the testing of the model and its association with management goals (Henriques & Pereira, 2024). Finally, deep learning has also been applied in the context of tourism-demand forecasting, which involves forecasting complex temporal patterns and the number of bookings (Dowlut & Gobin-Rahimbux, 2023).

Studies have shown that analysing reservations can lead to better predictions of reservations that are cancelled and better revenue planning. Machine-learning and big data techniques can detect high-risk bookings more efficiently than basic historical averages can and can help to facilitate proactive intervention (Sánchez-Medina & Eleazar, 2020). Also, regression-based machine-learning techniques can be used to enhance hotel-demand forecasts for pricing and capacity decisions (Pereira & Cerqueira, 2022). However, the accuracy of the forecast is not enough. Occupancy predictions rely on human judgment as well as combinations of models and the forecasting horizon, and analytical systems should be used in conjunction with, not instead of, managerial expertise (Schwartz et al., 2021). In addition, field evidence suggests that reinforcement-learning methods can enhance revenue-management decisions by fine-tuning pricing and capacity management in response to fluctuations in market demand (Chen et al., 2023).

Pricing behaviour is closely linked with cancellation risk. Last-minute practices in the hotel industry often involve dynamic pricing due to anticipated demand and the goal of bringing down the price of unsold rooms following cancellations, among other reasons (Guizzardi et al., 2024). These flexible conditions can affect customer behavior and flexibility to cancel and rebook, and dynamic adjustments can play a role in that. While prediction and forecasting have made progress, there is a need for more explanatory analysis, which seeks to elucidate the independent contribution of reservation characteristics that is not already captured in existing studies. Evidence that focuses on the determinant is important because it is important for managers to have interpretable information, not only in terms of which factors increase or decrease the likelihood of cancellation, but also in terms of the actual scores.

Thus, in this research, the study focuses on the analytics of the hotel booking cancellations, highlighting the hotel's history, booking time, business situation, price, channel structure and hotel features. The study combines descriptive, bivariate and multivariable evidence to understand the joint effect of the attributes of the reservations on cancellation behavior and what implications are relevant for revenue protection, risk-based policies and reliable operational planning.

Research Objectives

1. To analyze how many hotel bookings are cancelled and by how many categories of hotel reservation and customers.
2. To determine the customer, booking, pricing, channel and hotel-related factors that play a significant role in determining the likelihood of cancellation.
3. To assess how well a logistic regression model for the booking cancellation risk explains and predicts.

2. Methodology

2.1 Research Design

The study had a quantitative and explanatory research design used to find out the factors that influence hotel booking cancellations. The secondary data was used to investigate the influences of the following on the rate of customer cancellation: the customer characteristics, the booking characteristics, the pricing characteristics, the distribution characteristics, and the hotel characteristics.

2.2 Data Source

This study adopted the Hotel Booking Reservation dataset, consisting of 119,390 records of reservations with 33 variables (Bedmutha, 2025). Data collected covered booking status, lead time, type of hotel, customer attributes, market segment, type of deposit, average daily rate, room allocation, and previous booking behavior.

2.3 Study Population and Sample

All the hotel bookings in the data set comprised the study population, and each booking was considered a unit of analysis. The total number of cancelled bookings was 44224, and 75166 were non-cancelled bookings, giving an overall cancellation rate of 37.04%.

2.4 Study Variables

The dependent variable was cancelled (coded as 1 for cancelled bookings and 0 for non-cancelled bookings). The independent variables were lead time, hotel type, month of arrival, length of stay, number of guests, market segment, distribution channel, type of deposit, customer type, previous cancellations, booking changes, average daily rate, parking requirements, special requests, room type and city.

2.5 Data Preparation

Duplication, missing values, inconsistent data and extreme data were checked in the dataset. Records that had invalid guest counts or zero-night stays or negative average daily rates were reviewed and excluded as were records that had significant missing data. Items that had significant numbers of missing values were recoded or excluded from the primary analysis.

2.6 Derived Variables

Total length of stay was determined by summing up the number of nights spent during the week and over the weekend. The number of guests was arrived at by adding up the number of adults, children and babies. A variable was added to the room-mismatch, such as whether the assigned room type was different from the reserved room type.

2.7 Descriptive and Bivariate Analysis

Data was summarised using frequencies, percentages, means, standard deviations, medians and ranges. Chi-square tests were used for categorical variables, whereas independent-samples t-tests or Mann–Whitney U tests were used for continuous variables, depending on the distribution of the data.

2.8 Logistic Regression and Model Evaluation

Significant determinants of booking cancellation were determined by binary logistic regression. Regression coefficients, adjusted OR, 95% CI, and p-value were used for reporting results. The multicollinearity was examined via the VIF and tolerance values, and the model fit was evaluated via the Nagelkerke (R2), the Hosmer–Lemeshow test, classification accuracy, ROC curve and area under the curve.

2.9 Ethical Considerations

All the data used in this study was anonymised and unidentifiable. To avoid target leakage, the variables "reservation_status" and "reservation_status_date" were not included in the analytical model. All the findings were reported in aggregate, and statistical significance was set at $p<0.05$.

3. Results

3.1 Sample Characteristics and Cancellation Distribution

After cleaning, there were 118,564 hotel reservations. The total number of cancelled bookings was 44,176, representing 37.26% of the overall bookings. The number of completed or retained bookings was 74,388. The observations were made in city hotels (66.55%) and resort hotels (33.45%). Table 1 shows the profile of the hotel booking reservations.

Table 1. Profile of the Hotel Booking Reservations

Reservation characteristic	Frequency	Percentage (%)
Total reservations	118,564	100.00
Cancelled bookings	44,176	37.26
Non-cancelled bookings	74,388	62.74
City hotel bookings	78,899	66.55
Resort hotel bookings	39,665	33.45

A distinct difference in cancellation behaviour was noted between the two types of hotels. The cancellation rate for city hotels was 41.91%, which is higher than that of resort hotels, which was 28.01%, as shown in Figure 1. This preliminary result suggests that there was a significant association between the type of hotel and outcomes of hotel reservations

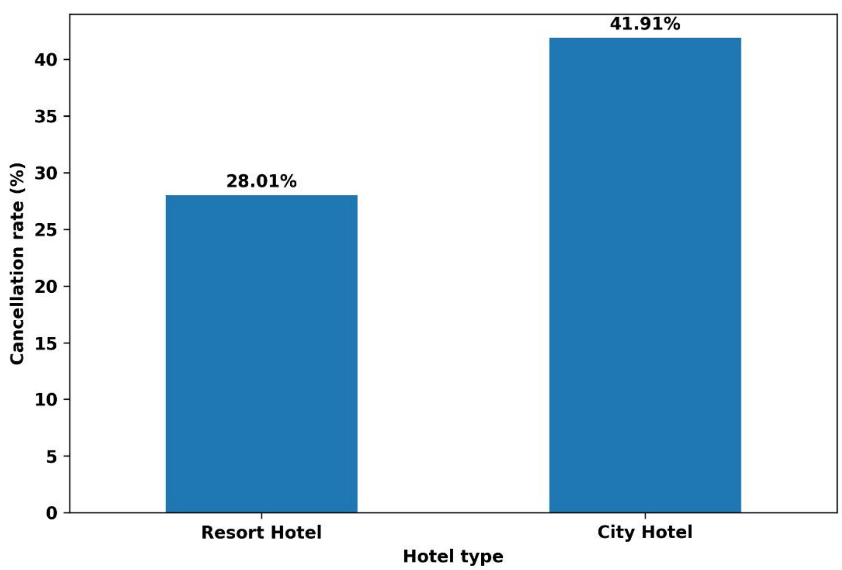

Figure 1. Booking Cancellation Rate by Hotel Type

3.2 Comparison of Continuous Reservation Characteristics

The continuous characteristics of cancelled and non-cancelled reservations are compared in Table 2. The average lead time for cancelled bookings was significantly longer at 144.95 days, whereas it was 80.49 days for non-cancelled bookings. The difference was significant ($p<0.001$) ($t=-98.28$).

The average daily rate for cancelled bookings was also slightly higher, and there were more guests and longer stays. The number of previous cancellations was significantly higher for cancelled bookings. Conversely, the number of booking changes and special requests was higher among non-cancelled bookings as compared to cancelled bookings.

Table 2. Comparison of Continuous Variables by Cancellation Status

Variable	**Overall mean ± SD**	**Non-cancelled mean**	**Cancelled mean**	**t-value**	**p-value**
Lead time, days	104.51 ± 106.92	80.49	144.95	−98.28	<0.001
Total stay, nights	3.44 ± 2.53	3.42	3.49	−4.68	<0.001
Total guests	1.97 ± 0.72	1.95	2.01	−14.55	<0.001
Average daily rate	102.52 ± 50.00	101.01	105.08	−13.29	<0.001
Previous cancellations	0.09 ± 0.85	0.02	0.21	−30.01	<0.001
Booking changes	0.22 ± 0.64	0.29	0.10	56.44	<0.001
Special requests	0.57 ± 0.79	0.72	0.33	89.11	<0.001

This relationship was especially strong for booking time and cancellation. Looking at Figure 2, the cancellation rate went up as time went up, with the highest rate found at bookings made more than 365 days ahead at 67.72%. With regard to Figure 2, the higher the number of days in advance the booking, the higher the cancellation rate, for instance, bookings made more than 365 days ahead of arrival were cancelled at a rate of 67.72%. In the pattern, the length of the booking period was correlated with greater uncertainty and cancellation risk.

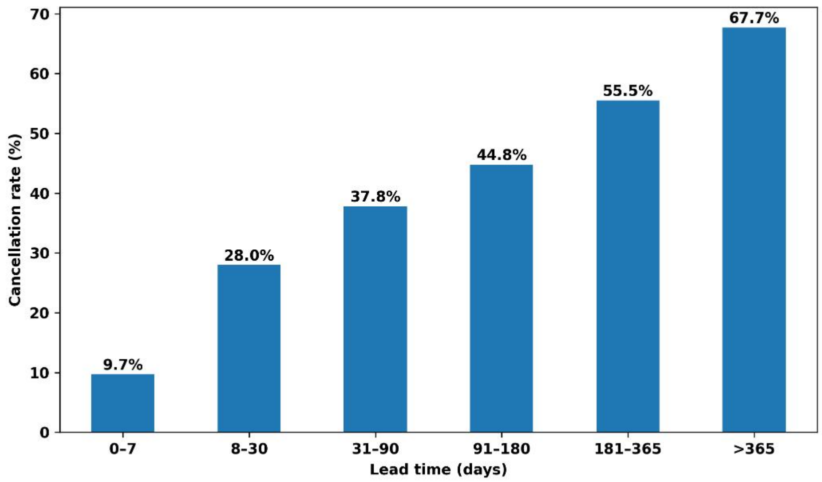

Figure 2. Cancellation Rate by Booking Lead Time

3.3 Cancellation Patterns Across Booking Categories

The hotel type, deposit policy, market segment, distribution channel, and the type of customer and whether they were a previous customer or not had significant differences in cancellation rates. There is a higher resort hotel booking cancellation rate than city hotel booking cancellation rate as reported in Table 3. The non-refundable reservations had an extremely high cancellation rate of 99.36% compared to 28.56% for bookings without a deposit and 22.22% for refundable reservations. When it comes to large market category cancellations, group bookings were the highest with 61.21%, followed closely by bookings through online travel agencies (OTAs) at 36.92%. The cancellation rate for direct bookings was much lower at 15.50%. Likewise, reservations through travel agents or tour operators had a cancellation rate of 41.22%, while direct channels resulted in a cancellation rate of 17.65%. The cancellation rate of transient customers was 41.01%, while the lowest one was of group customers, at 10.19%. The 37.91% cancelation rate for first-time guests was also a significant difference from the 15.69% rate for repeat guests.

Table 3. Cancellation Rates Across Major Reservation Categories

Variable	Category	Number of bookings	Cancelled bookings	Cancellation rate (%)
Hotel type	City Hotel	78,899	33,066	41.91
	Resort Hotel	39,665	11,110	28.01
Deposit type	Non Refund	14,587	14,494	99.36
	No Deposit	103,815	29,646	28.56
	Refundable	162	36	22.22
Market segment	Groups	19,758	12,093	61.21
	Online TA	56,129	20,723	36.92

	Offline TA/TO	24,049	8,301	34.52
	Corporate	5,231	989	18.91
	Direct	12,449	1,930	15.50
Distribution channel	TA/TO	97,335	40,123	41.22
	Corporate	6,584	1,462	22.21
	Direct	14,450	2,550	17.65
Customer type	Transient	88,935	36,473	41.01
	Contract	4,056	1,262	31.11
	Transient-Party	25,004	6,383	25.53
	Group	569	58	10.19
Repeated guest	No	115,066	43,627	37.91
	Yes	3,498	549	15.69
Room mismatch	No mismatch	104,110	43,383	41.67
	Mismatch	14,454	793	5.49

In Figure 3, the large variation between deposit policies is presented graphically. The data set contained an almost total cancellation incidence for non-refundable bookings. This is an unusual pattern, suggesting that deposit policy was among the most significant reservation-level determinants, but the result should be taken with a pinch of salt, as it might also be the result of the operational coding or booking channel of the data set.

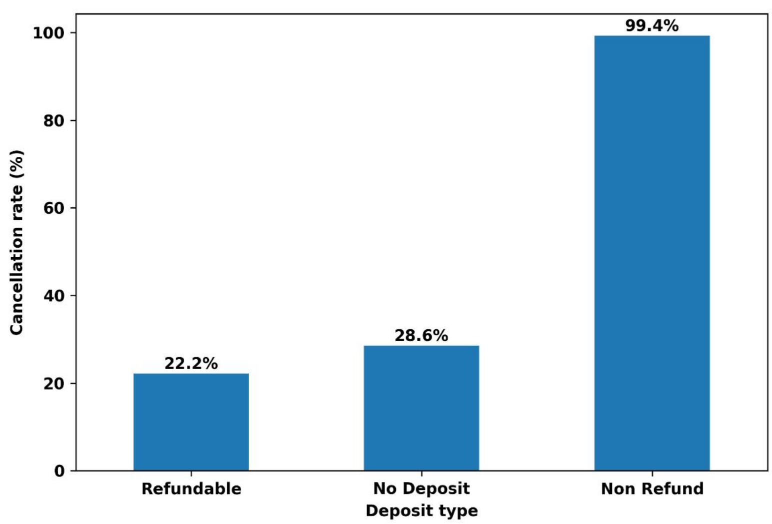

Figure 3. Cancellation Rate by Deposit Type

3.4 Seasonal Distribution of Booking Cancellations

Cancellation rates were fairly consistent throughout the months of arrival. The monthly rate from March to November was between 36.73% and 38.18%, as shown in Figure 4. While the differences may not have been significant from month to month, November had the highest cancellation rate, and March had the lowest.

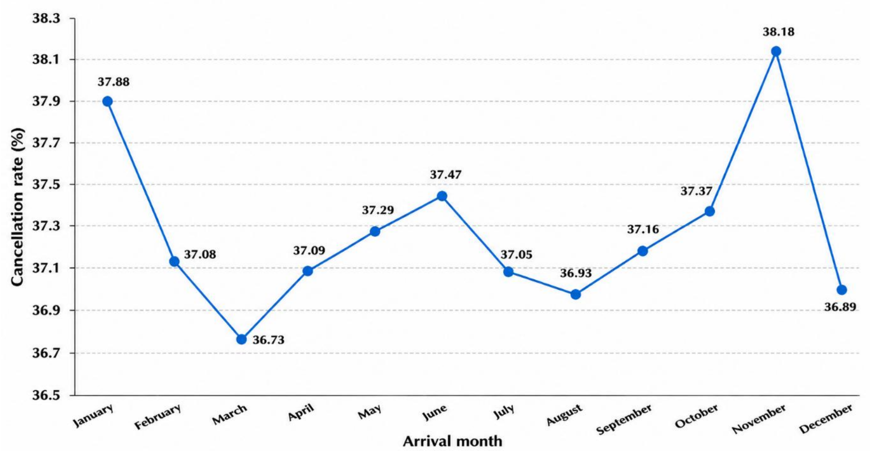

Figure 4. Monthly Variation in Booking Cancellations

This small difference month by month was attributed to the fact that the characteristics of the reservations themselves (lead time, cancellation history, deposit policy and type of customer) were more important than the month of the arrival.

3.5 Determinants of Hotel Booking Cancellation

A binary logistic regression was performed to examine the adjusted effects of the reservation characteristics on the probability of cancellation. The adjusted ORs for continuous predictors were based upon standardisation before estimation; thus, they reflect the effect of an increase in the predictor of one standard deviation.

As shown in Table 4, The longer the lead time, the more likely the cancellation. An increase in lead time by 1 standard deviation was associated with an adjusted odds ratio of 1.491 ($p<0.001$) and an increase in the odds of cancellation. There was also a correlation between increased cancellation probability and longer stays, greater number of guests, and greater mean daily room charge.

One of the most powerful customer-level predictors was previous cancellation behaviour. The adjusted odds ratio (9.432) was greater than 1, and the trend was significant ($p<0.001$) when the odds of cancellation increased by more than nine times with a 1 standard deviation increase in previous cancellations. By comparison, cancelled bookings were reduced by ~52.26% due to previous non-cancelled bookings.

There was a negative relationship between booking changes, special requests and cancellations. The more the standard deviation of booking changes, the lower the odds of cancellation (about

21.03% less for each one standard deviation of booking change) and the more the standard deviation of special requests, the lower the odds of cancellation (about 43.23% less for each one standard deviation of special request). The findings indicate that the higher the interaction with the reservation, the more customers feel committed to the reservation.

Table 4. Binary Logistic Regression Predicting Hotel Booking Cancellation

Predictor	B	Adjusted OR	95% CI for OR	p-value
Lead time	0.400	1.491	1.463–1.520	<0.001
Total stay	0.105	1.110	1.093–1.128	<0.001
Total guests	0.073	1.075	1.057–1.094	<0.001
Average daily rate	0.121	1.129	1.109–1.149	<0.001
Previous cancellations	2.244	9.432	8.557–10.397	<0.001
Previous non-cancelled bookings	−0.739	0.477	0.443–0.514	<0.001
Booking changes	−0.236	0.790	0.775–0.805	<0.001
Days on waiting list	−0.006	0.994	0.978–1.011	0.482
Special requests	−0.566	0.568	0.558–0.578	<0.001
Resort Hotel[1]	−0.069	0.934	0.902–0.966	<0.001
Non Refund deposit[2]	5.462	235.623	189.069–293.641	<0.001
Refundable deposit[2]	0.155	1.168	0.774–1.763	0.460
Online TA market segment[3]	0.855	2.352	1.638–3.378	<0.001
TA/TO distribution channel[4]	−0.114	0.893	0.778–1.024	0.106
Transient customer[5]	0.839	2.313	2.085–2.566	<0.001
Repeated guest	−0.575	0.563	0.475–0.666	<0.001
Room mismatch	−1.810	0.164	0.151–0.177	<0.001

Resort hotel bookings were about 6.63% less likely to be cancelled after adjustment for other predictors. The probability of the OTA reservations being cancelled was over 2 times that of the reference market segment. The odds of cancellation were 2.31 times more likely for transient customers than for contract customers.

As a result, the odds of repeat visitors cancelling were roughly 43.72% less than the odds of first-time visitors cancelling. The odds of cancellation were significantly lower for room mismatch, too. This correlation must be taken with caution, however, as assignment of rooms may be done after the flight is booked, and thus could be reflective of operational actions near the time of flight. After adjusting for the other reservation factors, days on the waiting list and the distribution channel were not statistically significant.

3.6 Model Performance

The logistic regression model had fair explanatory and discriminatory ability. The McFadden pseudo-(R^2) was 0.338, indicating that the sole intercept model was much less adequate in terms of the characteristics of reservations that were included. The sum of the area under the curve for the entire sample was 0.850.

The model had an overall accuracy of 81.62% with a precision of 83.73%, sensitivity of 62.89%, specificity of 92.74% and an F1 score of 71.82% in the independent test sample. The test AUC was good, 0.866, discriminating between cancelled and non-cancelled reservations. Table 5 shows the Logistic Regression Model Fit and Predictive Performance.

Table 5. Logistic Regression Model Fit and Predictive Performance

Model statistic	**Value**
Number of observations	118,559
Log likelihood	−51,812.291
McFadden pseudo-(R^2)	0.338
Akaike information criterion	103,700.582
Bayesian information criterion	104,068.542
Full-sample AUC	0.850
Test accuracy	81.62%
Test precision	83.73%
Test sensitivity	62.89%
Test specificity	92.74%
Test F1-score	71.82%
Test AUC	0.866
True negatives	13,798
False positives	1,080
False negatives	3,279
True positives	5,556

The receiver operating characteristic curve (ROC) in Figure 5 continued to be substantially higher than the chance diagonal. AUC of 0.866 means that the model had 86.6% chance of giving a higher score for cancellation risk to a randomly selected cancelled booking than to a randomly selected booking that was not cancelled.

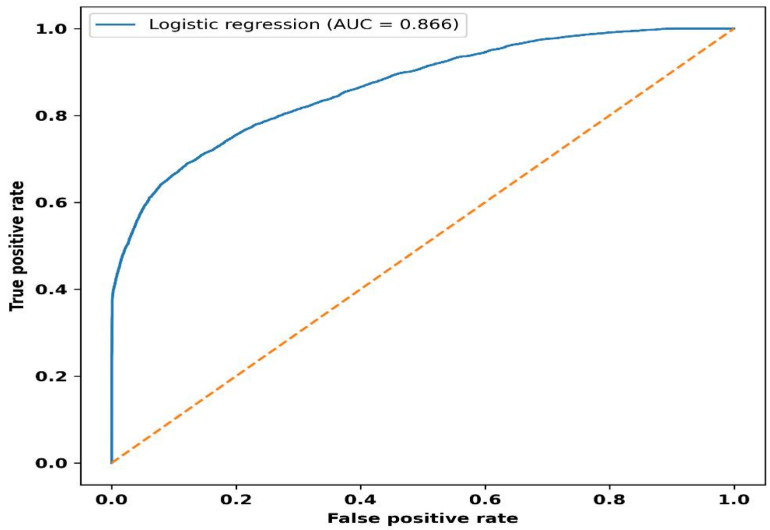

Figure 5. Receiver Operating Characteristic Curve for the Cancellation Model

3.7 Summary of Major Findings

The findings revealed that the significant positive factors that determined the hotel booking cancel were booking lead time, previous hotel booking cancellations, type of deposits, market segment, customer type, average daily rate, length of stay, and party size. Previous non-cancelled bookings, booking modification, special requests, repeated guest status and resort hotel classification all had a significant impact on decreasing the probability of cancellation. Non-refundable deposit status and previous cancellation history were the strongest positive relationships with cancellation, among all variables. In general, the results indicate that the analytics on reservations can be used effectively to identify high-risk bookings and to predict cancellation and non-cancellation of reservations with a relatively high degree of accuracy.

4. Discussion

The present study found that there are some factors in the reservations and customer, pricing and the hotel, which significantly affect the booking cancellation behaviour. The results of this study show that hotel cancellations are not a random occurrence but have measurable properties like lead time, previous cancellation, deposit policy, market segment, type of customer, modification of the booking, special requests and hotel category. This contributes to the general adoption of reservation analytics as a viable means to predict cancellation risk and make better RM decisions.

Lead time was a significant and positive predictor of cancellation. In general, the cancellation rate was significantly higher for reservations made far in advance than for reservations made close to the date of arrival. This is reasonable because the longer the booking window, the more time customers have to alter their travel plans, consider options, and adjust to price and personal changes. In other studies, the booking time is found to be one of the most influential factors affecting booking-cancellation classification and booking forecasting models (Andriawan et al., 2020; Herrera et al., 2024). Two-stage predictive methods have also been found to be effective in identifying cancellations (Kundu et al., 2025) with better temporal booking characteristics.

One of the strongest effects in the logistic regression model was cancellation behaviour, which had previously been found. The likelihood of cancellation was significantly higher for guests who had cancelled previously, while past bookings not cancelled increased the likelihood of non-cancellation. This result highlights that the past behaviour of customers can be a valuable guide for the reliability of their bookings. Bayesian and comparative modelling studies also show that the estimation of the risk of cancellation is improved with the use of past reservation patterns (Jishan et al., 2024; Русакова & Радионова, 2021). Hotels might then use past booking results as part of customer-risk profiles, with the assurance that automated decisions are based fairly and that customers will not be unfairly penalised.

Another factor that had a significant impact was deposit type. The cancellation rate of non-refundable reservations was extremely high in the data analysed. This seems counter intuitive as it would suggest one would not expect to see non-refundable conditions and cancellation outcomes in such a relationship, but similar datasets have occasionally produced odd associations between the deposit category and booking results. Patterns can occur in channels, may be due to group bookings, operational coding or customers not wanting to incur a loss if travel plans change. The figures obtained should, therefore, be taken with a pinch of salt and checked against booking-policy documentation. Profit-driven cancellation studies highlight that the financial value of a predictor is more than just its statistical link; it is also driven by the monetary impact of misclassification and lost bookings (Liu et al., 2023, 2025).

It was found that online travel agencies (OTAs) and some intermediary booking methods had higher cancellation risk than direct booking. Online channels allow for more simple comparison and reservation switching to give customers flexible alternatives until they make a decision. This highlights the need for specific cancellation controls and confirmation measures per channel. Similarly, studies conducted based on classification have proven that market segment and distribution channel is a useful predictor of booking outcome (Chen et al., 2022; Lei & Ibrahim, 2024). Hotels can thus use different reminders and/or deposit terms/cancellation periods for different channels based on the risk profile of those channels.

The customer engagement indicators were found to be negatively correlated with cancellation. Those who returned were less likely to cancel the booking; so were guests with more special requests and those who booked for booking changes. These qualities can be associated with a higher level of commitment, hotel engagement or clearer travel aspirations. The added value of being a regular guest also implies that booking stability is enhanced when booking with regular customers (Liu et al., 2025). In recent years, machine learning research has shown that features such as special requests, booking changes, and customer history are all informative in the context of predicting booking cancellations (Bhardwaj et al., 2024; Rahmawati et al., 2024). The following variables could help the hotels differentiate between real customers and 'speculative' customers.

The model achieved a good discriminatory ability, the AUC for the test set was 0.866, and the accuracy was above 81%. This suggests that the chosen predictors are useful to differentiate between cancelled and non-cancelled bookings. Similar investigations revealed that machine-learning techniques attain good predictive accuracy, especially if nonlinear interactions and class distribution are handled properly (Adil et al., 2021; Hermawan et al., 2025). Predictive approaches can also be enhanced by adaptive methods, which can be more useful in the case of shifting customer behavior and changing market conditions over time (Silvestre et al., 2026). However, the usefulness of the model doesn't depend on accuracy. Other factors, such as precision, sensitivity, cost of operation, and revenue effect, should also be considered.

The results have managerial implications. The score of a cancellation risk can be used for prioritizing confirmation messages, for changing the overbooking percentage, for implementing different deposit levels, and for preempting high-risk bookings before they are confirmed. But models should be interpretable; they have to be combined with managerial judgment. Explanable and profit-oriented frameworks are particularly useful as they link forecasts to action in the business and financial implications.

5. Conclusion

This study analyzed the causes of cancelled hotel bookings by analyzing the hotel booking data and found a few factors that affected the booking cancellation behaviour of the users, which were categorized in three distinct areas: Customer, Booking, Pricing, and Hotel. The results indicated that the longer the lead time, the more likely it was to be cancelled, as well as being a non-refundable deposit, having a higher average daily rate, greater guest group size, and selected market segments. However, risk of cancellation was found to be negatively associated with repeated guest status, previous non-cancelled bookings, resort hotel classification, booking modifications and special requests. Overall, the logistic regression model performed well both in terms of explanatory and predictive capability, as its area under the curve was 0.866 and the overall accuracy was more than 81%. The results corroborate the feasibility of using the information at the reservation level to effectively differentiate high-risk bookings from more stable reservations. The analysis also pointed to the need to interpret unusual patterns with caution, especially the close link between non-refundable bookings and booking cancellations, as well as to consider operational coding and booking-channel practices when interpreting these patterns. At the management level, hotels can use cancellation-risk estimates to enhance their demand forecasting, guests policies on accepting deposits, and inventory allocation. For long lead-time bookings and customers who have canceled bookings before, targeted reminders and confirmation requests can be particularly helpful.

Reference

1. Adil, M., Ansari, M. F., Alahmadi, A., Wu, J. Z., & Chakrabortty, R. K. (2021). Solving the problem of class imbalance in the prediction of hotel cancelations: A hybridized machine learning approach. Processes, 9(10), 1713.

2. Altin, M., Chen, C. C., Riasi, A., & Schwartz, Z. (2023). Go moderate! How hotels' cancellation policies affect their financial performance. Tourism Economics, 29(8), 2165-2182.

3. Andriawan, Z. A., Purnama, S. R., Darmawan, A. S., Wibowo, A., Sugiharto, A., & Wijayanto, F. (2020, November). Prediction of hotel booking cancellation using CRISP-DM. In 2020 4th International Conference on Informatics and Computational Sciences (ICICoS) (pp. 1-6). IEEE.

4. Atabay, L., Çizel, B., & Atabay, E. (2026). Timing is everything: Segmenting OTA's cancellations behavior through the BCR model. International Journal of Hospitality Management, 135, 104605.

5. Bedmutha, K. S. (2025). Hotel booking reservation: An updated 2024 dataset for hotel booking demand and cancellation analysis [Data set]. Kaggle. https://www.kaggle.com/datasets/kundanbedmutha/hotel-booking-reservation

6. Bhardwaj, A., Yadav, T., & Chaudhary, R. (2024, June). Predicting hotel booking cancellations using machine learning techniques. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

7. Boto-García, D., Zapico, E., Escalonilla, M., & Pino, J. F. B. (2021). Tourists' preferences for hotel booking. International Journal of Hospitality Management, 92, 102726.

8. Chen, J., Xu, Y., Yu, P., & Zhang, J. (2023). A reinforcement learning approach for hotel revenue management with evidence from field experiments. Journal of Operations Management, 69(7), 1176-

1201.

9. Chen, Y., Ding, C., Ye, H., & Zhou, Y. (2022, March). Comparison and analysis of machine learning models to predict hotel booking cancellation. In 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022) (pp. 1363-1370). Atlantis Press.

10. Das, R., Chadha, H., & Banerjee, S. (2023). Multi-layered market forecast framework for hotel revenue management by continuously learning market dynamics. In Artificial Intelligence and Machine Learning in the Travel Industry: Simplifying Complex Decision Making (pp. 145-161). Cham: Springer Nature Switzerland.

11. Dowlut, N., & Gobin-Rahimbux, B. (2023). Forecasting resort hotel tourism demand using deep learning techniques–A systematic literature review. Heliyon, 9(7).

12. Guizzardi, A., Ballestra, L. V., & D'Innocenzo, E. (2024). Reverse engineering the last-minute on-line pricing practices: an application to hotels: A. Guizzardi et al. Statistical Methods & Applications, 33(3), 943-971.

13. Henriques, H., & Pereira, L. N. (2024). Hotel demand forecasting models and methods using artificial intelligence: A systematic literature review. Tourism & Management Studies, 20(3), 39-51.

14. Hermawan, A., Saputra, A., Lailinajma, N., Julianti, R., Hartanto, T., & Daniel, T. K. (2025). Predicting Hotel Booking Cancellations Using Machine Learning for Revenue Optimization. Router: Jurnal Teknik Informatika Dan Terapan, 3(1), 37-48.

15. Herrera, A., Arroyo, Á., Jiménez, A., & Herrero, Á. (2024). Forecasting hotel cancellations through machine learning. Expert Systems, 41(9), e13608.

16. Jishan, M. A., Singh, V., Ghosh, A. K., Alam, M. S., Mahmud, K. R., & Paul, B. (2024, December). Hotel Booking Cancellation Prediction Using Applied Bayesian Models. In 2024 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 1-5). IEEE.

17. Kmiecik, M., & Antonio, N. (2026). Cancel, rebook, save: Revenue leakage from price cuts in hotels. International Journal of Hospitality Management, 135, 104647.

18. Kundu, S., Roy, S., Shukla, A., & Mateen, A. (2025). Unveiling cancellation dynamics: A two-stage model for predictive analytics. Data & Knowledge Engineering, 160, 102467.

19. Lei, C., & Ibrahim, M. (2024). Efficient revenue management: Classification model for hotel booking cancellation prediction. Global Perspectives in Management, 2(1), 12-21.

20. Liu, Z., De Bock, K. W., & Zhang, L. (2025). Explainable profit-driven hotel booking cancellation prediction based on heterogeneous stacking-based ensemble classification. European Journal of Operational Research, 321(1), 284-301.

21. Liu, Z., Jiang, P., Wang, J., Du, Z., Niu, X., & Zhang, L. (2023). Hospitality order cancellation prediction from a profit-driven perspective. International Journal of Contemporary Hospitality Management, 35(6), 2084-2112.

22. Masiero, L., Viglia, G., & Nieto-Garcia, M. (2020). Strategic consumer behavior in online hotel booking. Annals of Tourism Research, 83, 102947.

23. Mir, V. R., Tortella, B. D., Leoni, V., & Nicolau, J. L. (2026). Decoding booking cancellations: Quantitative insights and theoretical advances in tourism behavior. Tourism Management, 114, 105384.

24. Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. Current Issues in Tourism, 25(17), 2733-2750.

25. Rahmawati, E. K. A., Nurohim, G. S., Agustina, C., Irawan, D., & Muttaqin, Z. (2024). Develompent of Machine Learning Model to Predict Hotel Room Reservation Cancellations. Jurnal Teknologi Informasi dan Terapan, 11(2), 58-65.

26. Sánchez-Medina, A. J., & Eleazar, C. (2020). Using machine learning and big data for efficient forecasting of hotel booking cancellations. International Journal of Hospitality Management, 89, 102546.

27. Schwartz, Z., Webb, T., van der Rest, J. P. I., & Koupriouchina, L. (2021). Enhancing the accuracy of revenue management system forecasts: The impact of machine and human learning on the effectiveness of hotel occupancy forecast combinations across multiple forecasting horizons. Tourism Economics, 27(2), 273-291.

28. Silvestre, P., Antonio, N., & Carrasco, P. (2026). Navigating uncertainty: enhancing hotel cancellation predictions with adaptive machine learning. Information Technology & Tourism, 28(1), 9.

29. Русакова, Е. И., & Радионова, М. В. (2021). Прогнозирование отмены бронирования отелей: сравнительная характеристика спецификаций моделей. Вестник Пермского университета. Серия: Экономика, 16(4), 327-345.