★★★★★

Hotel Reservation Cancellation Analysis

Contents

1

Business Background and Objectives

Business Background

The typical reasons for cancellations include change of plans, scheduling conflicts, etc. This is often made easier by the option to do so free of charge or preferably at a low cost which is beneficial to hotel customers but it is a less desirable and possibly revenue-diminishing factor for hotels to deal with. Hotel reservation cancellations can have negative impacts on the hotel company including:

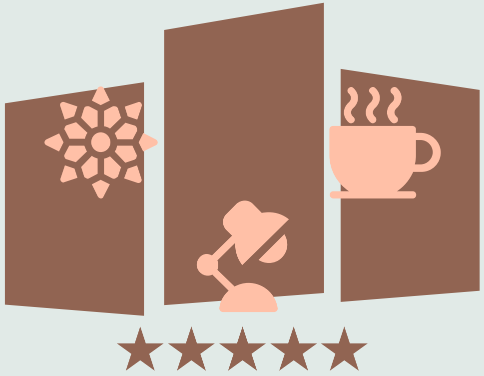
Loss of Revenue

Decrease in Occupancy

Impact on the hotel's image

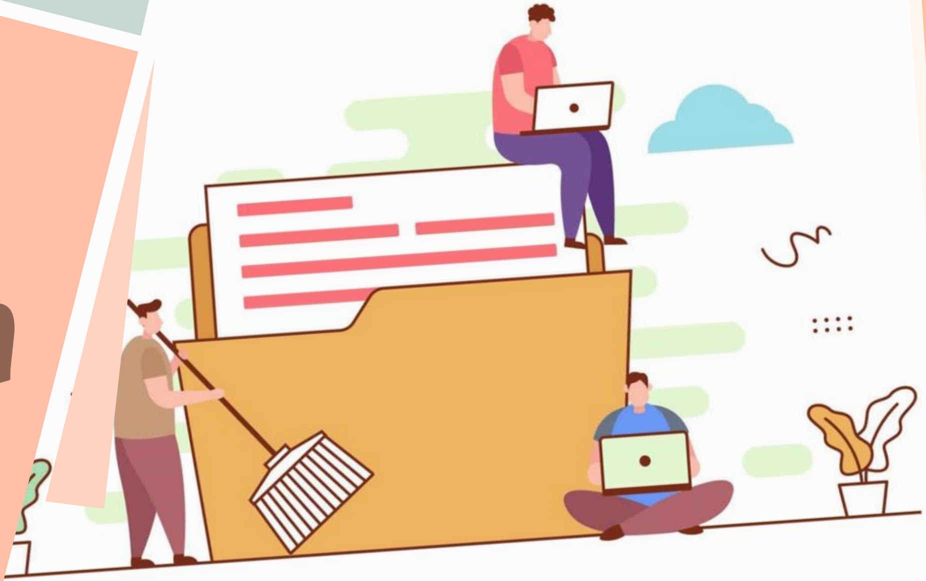
Objectives

- What kind of customer that has tendency to be canceled reservation?
- What are the influential features in reservation cancellations?
- What is the impact of the model to the business?



2

Data Preparation



Dataset Information

Booking_ID	Unique identifier of each booking
no_of_adults	Number of adults
no_of_children	Number of Children
no_of_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
no_of_week_nights	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
type_of_meal_plan	Type of meal plan booked by the customer
required_car_parking_space	Does the customer require a car parking space? (0 - No, 1- Yes)
room_type_reserved	Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels.
lead_time	Number of days between the date of booking and the arrival date
arrival_year	Year of arrival date
arrival_month	Month of arrival date
arrival_date	Date of the month
market_segment_type	Market segment designation.
repeated_guest	Is the customer a repeated guest? (0 - No, 1- Yes)
no_of_previous_cancellations	Number of previous bookings that were canceled by the customer prior to the current booking
no_of_previous_bookings_not_canceled	Number of previous bookings not canceled by the customer prior to the current booking
avg_price_per_room	Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
no_of_special_requests	Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
booking_status	Flag indicating if the booking was canceled or not.

Dataset Information

36,275
Rows

19
Features

1
Target
booking_status

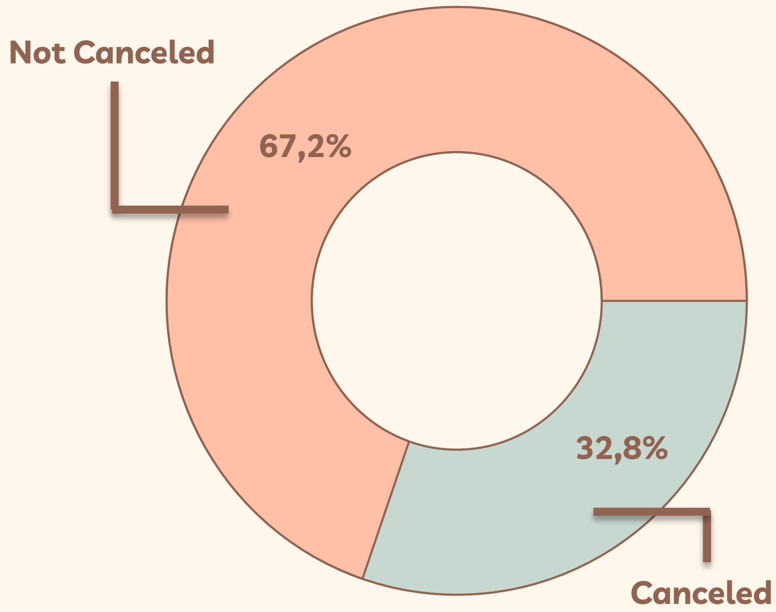
0
Missing Values

10,273
Duplicates
(After drop booking_ID)

3

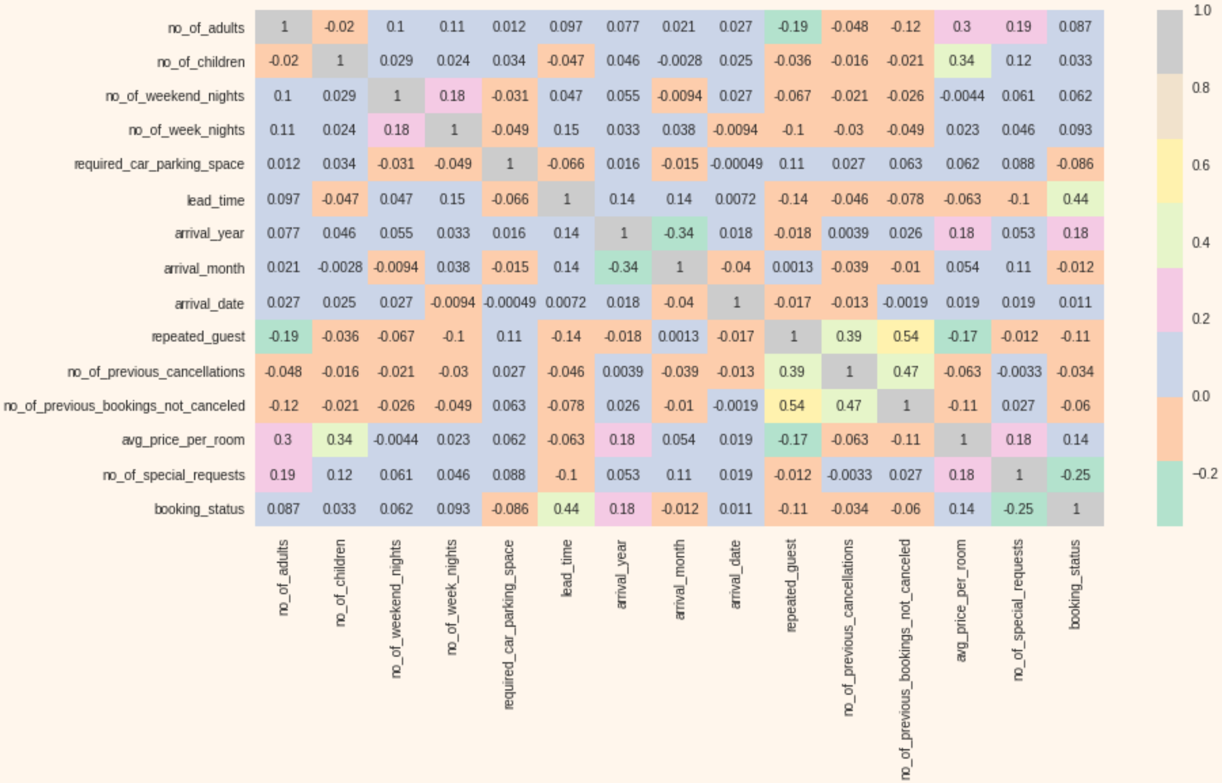
Exploratory Data Analysis

Comparison Target Variables



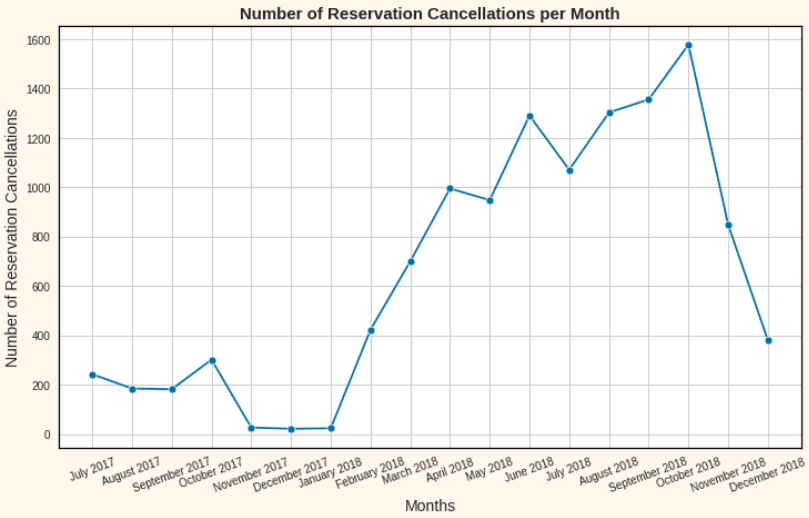
Approximately **32,8%** of the total hotel reservations were cancelled

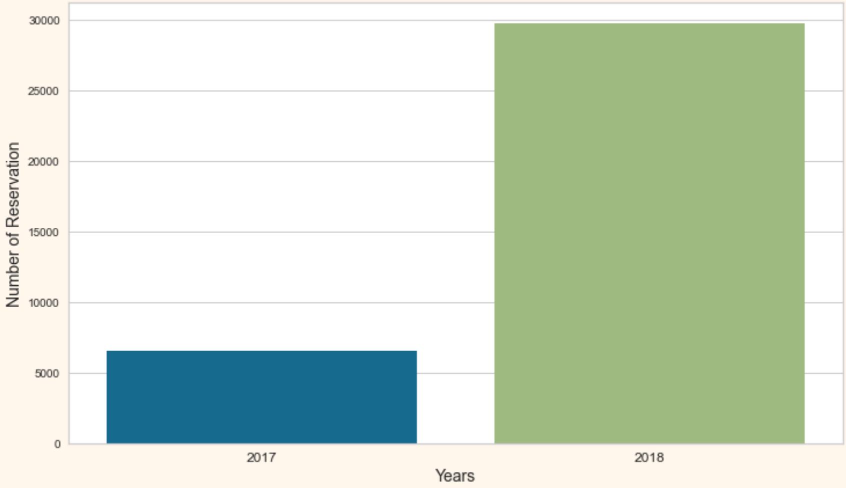
Multivariate Analysis



The features don't have strong correlation between the variables.

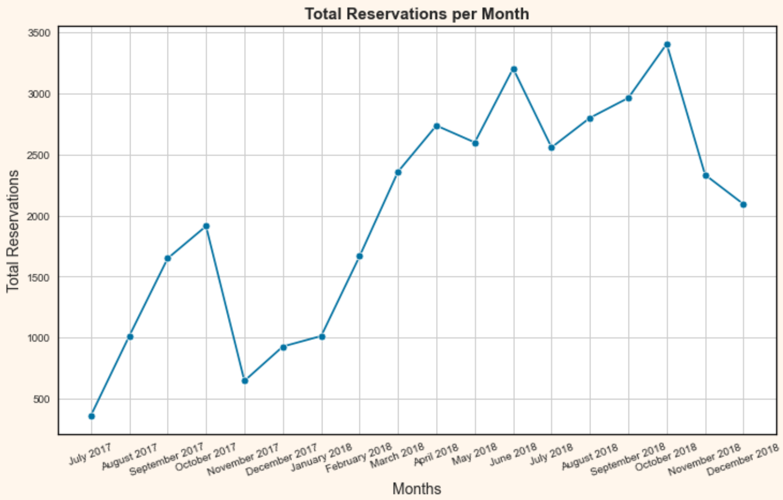
Time Based Analysis

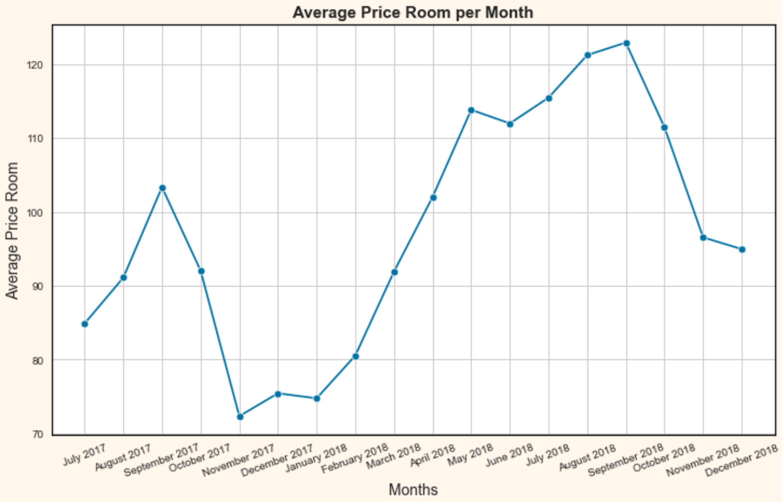




- After reaching its peak in October, the number of reservation cancellations immediately decreased in November and December. This could be due to the Christmas and New Year holidays, as many people go on vacation and rent hotel rooms during this time
- The number of reservation cancellations hugely increased from the end of 2017 to 2018, This is due to a significant increase in the number of reservations in 2018

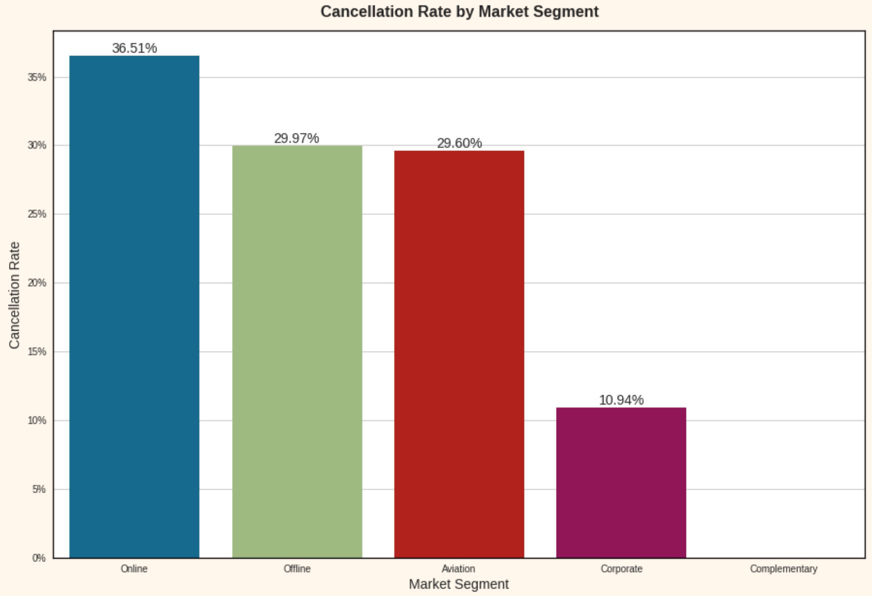
Time Based Analysis

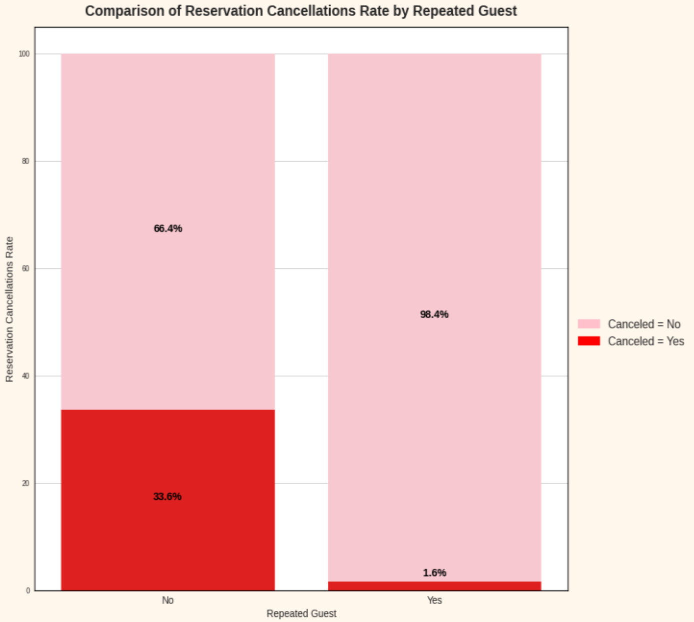




- The number of reservation cancellations reaches its peak every October in each year. This is because the number of reservations also peaks in October, but the average room rental prices reach their peak in September each year. It is recommended that the company increase the average room rental prices during peak demand periods and implement a cancellation policy during the low season with only partial refunds to minimize losses when customers cancel their reservations. However, during peak season, reservation cancellations are not charged because of high demand, and if there are cancellations, they can be quickly replaced by other customers

Categorical Features Analysis

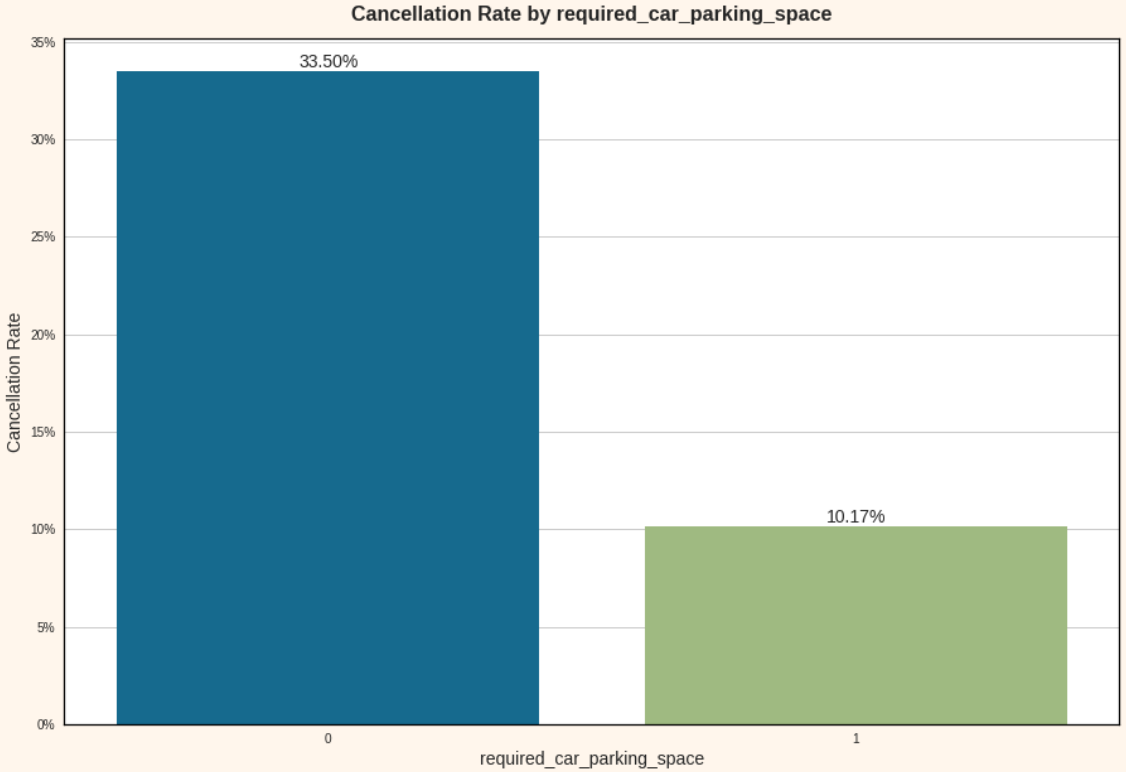




- Online reservations are more likely to be canceled with a probability of 36.51%
- Complementary reservations exhibit a low cancellation rate, with zero cancellations.

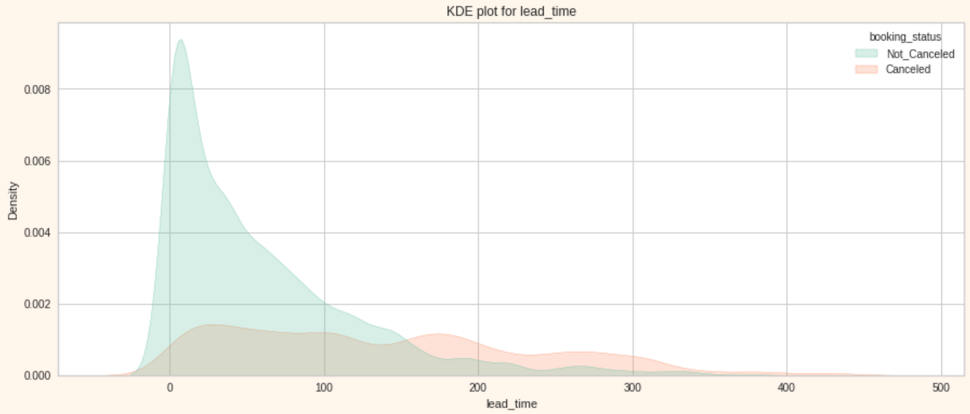
- Previous customers tend to be more loyal customers, with a very low reservation cancellation rate of 1.6%.
- New customers tends to be more likely to cancel their reservation easily, with reservation cancellation rate of 33.6%.

Categorical Features Analysis (Cont...)



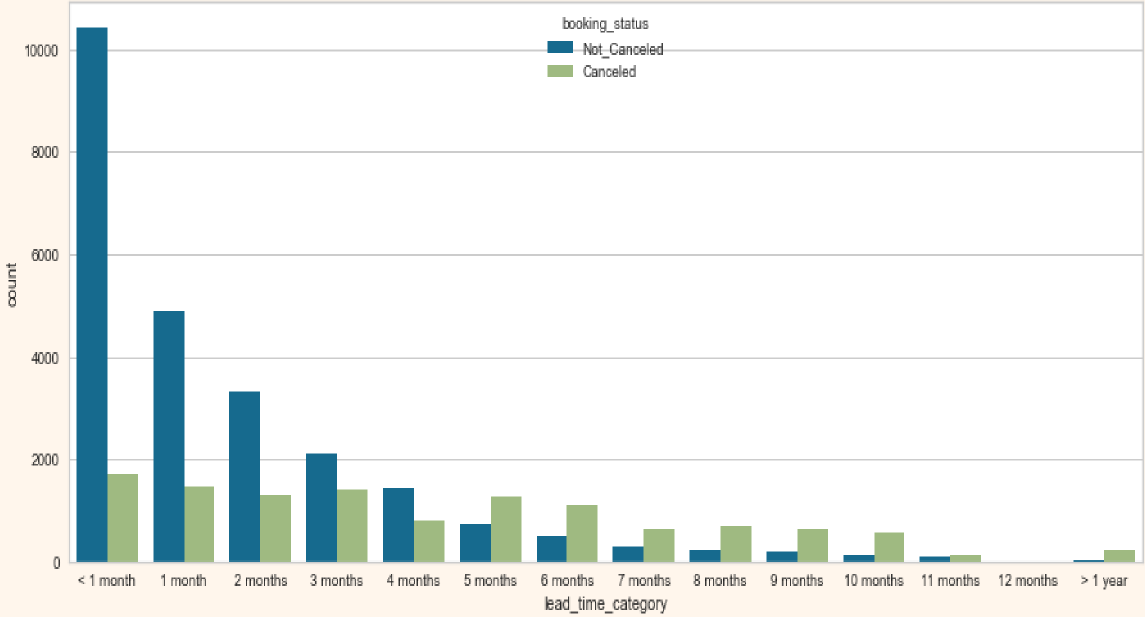
Customers who do not bring a vehicle are more likely to cancel their reservation.

Lead Time Analysis



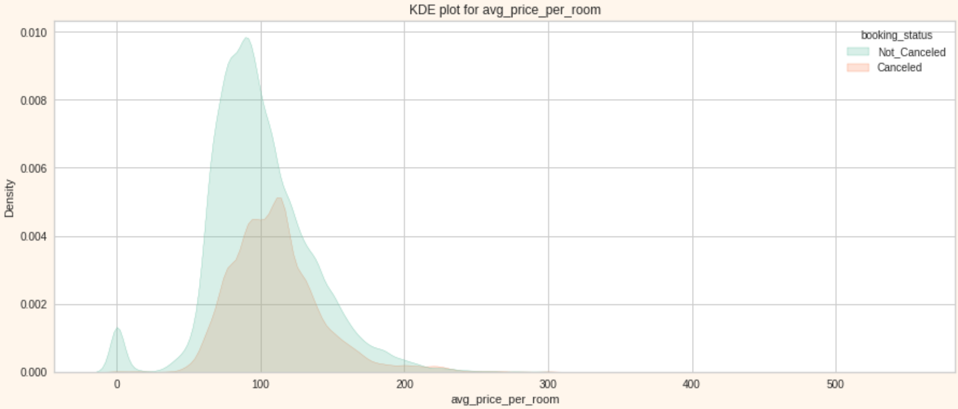
Customers with high lead times are more likely to cancel their reservations than those with low lead times, meaning that customers who book well in advance are more likely to cancel their reservations.

Lead Time Analysis



Consider shortening the maximum reservation time to one year from the current date to reduce cancellations for reservations made too far in advance. However, assess the potential impact on customer satisfaction and bookings before implementing the policy change.

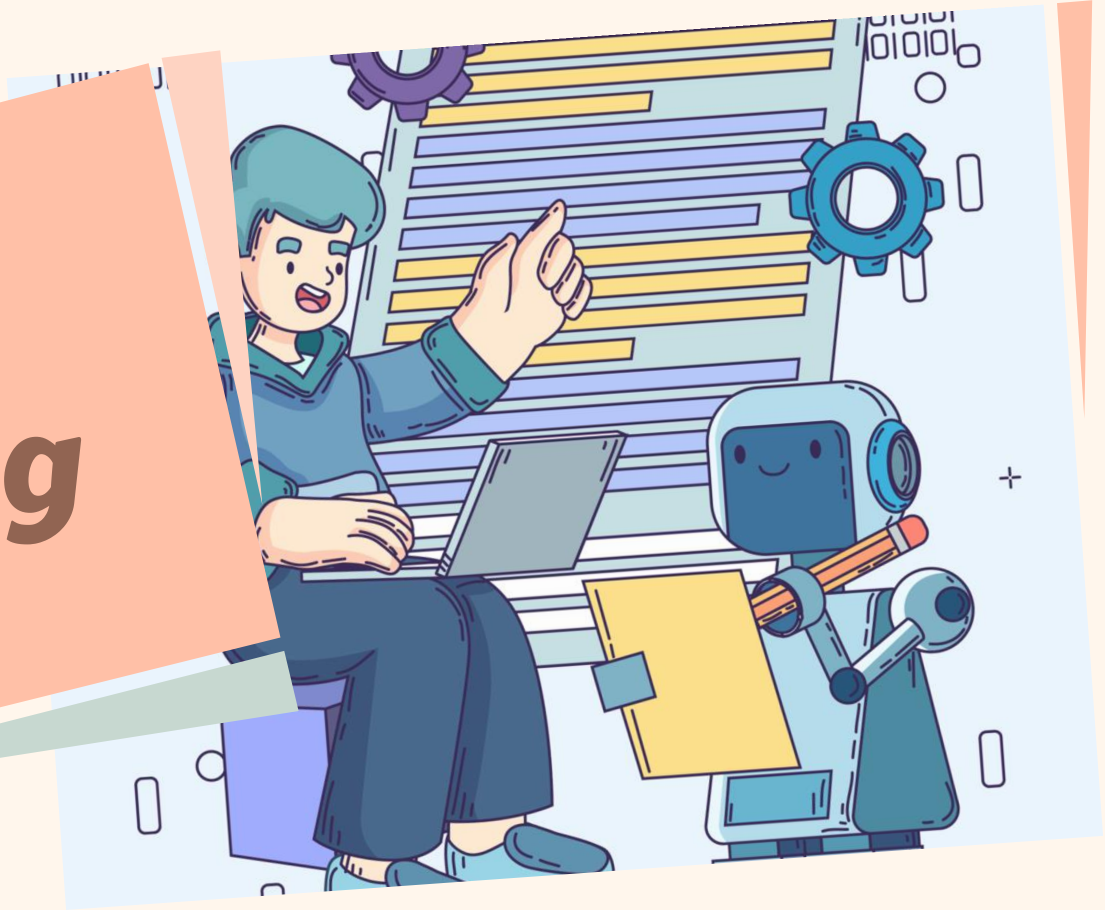
Average Price per Room Analysis



Customers who do not spend any money on room charges are less likely to cancel their reservations, and customers who spend between 100-120 euros on room charges are the peak of reservation cancellations

4

Modeling



Baseline Classifier Model

The dataset is splitted into train data (80%) and test data (20%)

Legend:

Accuracy: how accurate a model is in correctly classifying data

AUC: Measuring how well a model can distinguish between positive and negative classes, the closer the value is to 1, the better

Precision: Predicted positive rate

Recall: Actual positive rate

F1-Score: Combines precision and recall

Baseline Classifier Model

Training Data

Algorithms	AUC	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.687	0.778	0.654	0.474	0.550
KNN Classifier	0.762	0.834	0.772	0.594	0.671
Decision Tree	0.992	0.996	1	0.985	0.992
Random Forest	0.994	0.996	0.994	0.991	0.992
LightGBM	0.837	0.881	0.828	0.735	0.779
Gradient Boost	0.800	0.854	0.784	0.674	0.725

Baseline Classifier Model

Test Data

Algorithms	AUC	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.685	0.774	0.650	0.473	0.547
KNN Classifier	0.655	0.745	0.576	0.442	0.500
Decision Tree	0.773	0.813	0.677	0.676	0.676
Random Forest	0.800	0.852	0.779	0.679	0.726
LightGBM	0.810	0.860	0.795	0.693	0.741
Gradient Boost	0.792	0.845	0.767	0.666	0.713

We will choose model without significant difference between training data score and testing data score to avoid overfitting model

After Oversampling (Random Over Sampler) Classifier Model

★★★★★

Training Data

Algorithms	AUC	Accuracy	Precision	Recall	F1-Score
LightGBM	0.870	0.870	0.875	0.864	0.869
Gradient Boost	0.812	0.812	0.835	0.779	0.806

Test Data

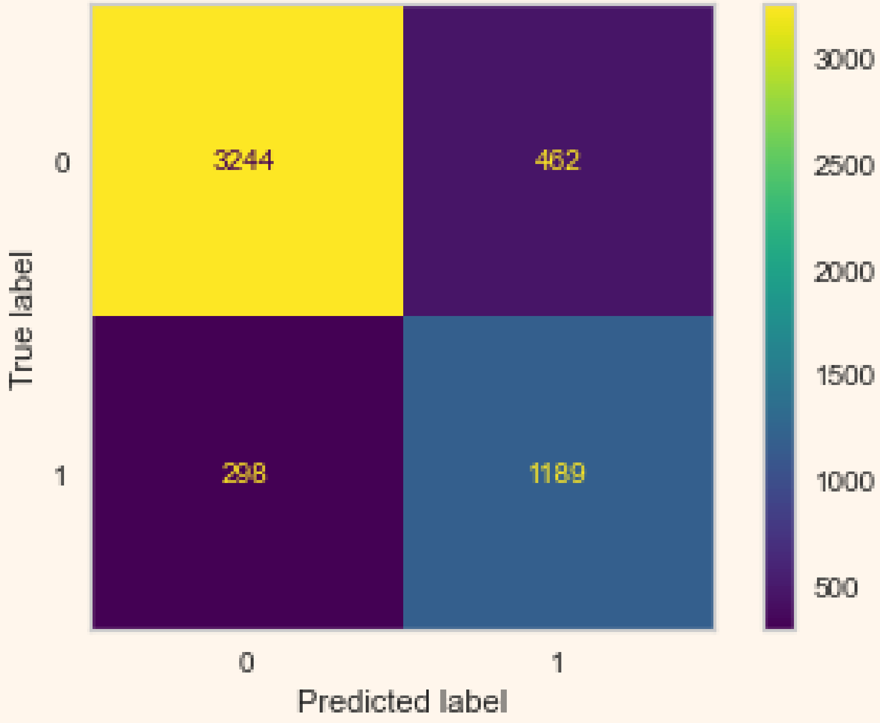
Algorithms	AUC	Accuracy	Precision	Recall	F1-Score
LightGBM	0.837	0.854	0.720	0.800	0.758
Gradient Boost	0.808	0.828	0.678	0.761	0.717

Hyperparameter Tuning Model

Algorithms	AUC	Accuracy	Precision	Recall	F1-Score
Tuned LightGBM	0.824	0.852	0.732	0.761	0.746
Base LightGBM	0.837	0.854	0.720	0.800	0.758

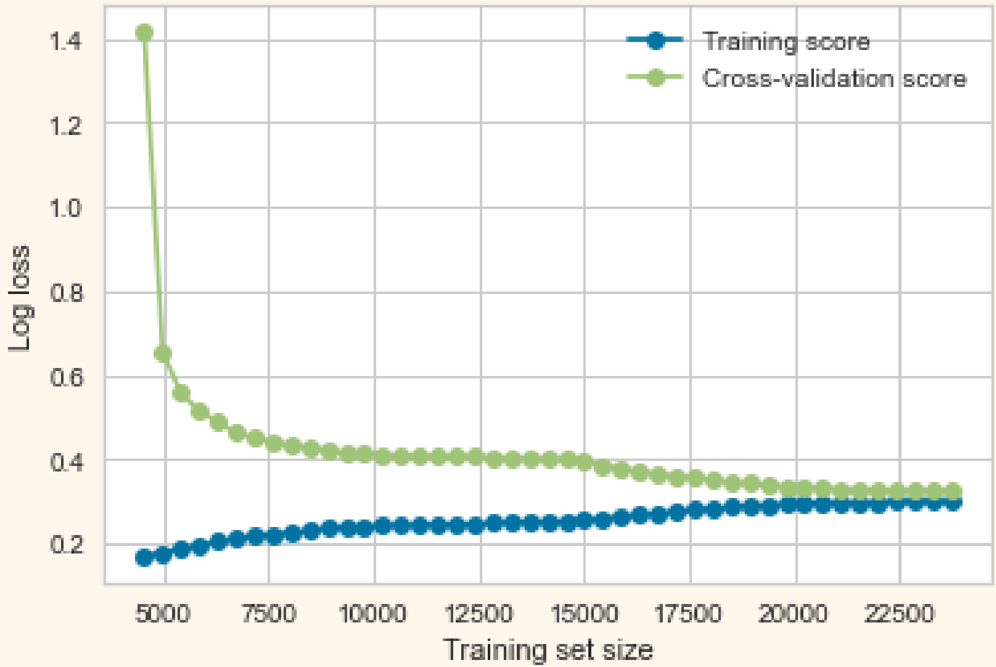
The selected model is the **LightGBM model without hyperparameter tuning and oversampling (Random Over Sampler)** applied to the dataset

Metrics Evaluation



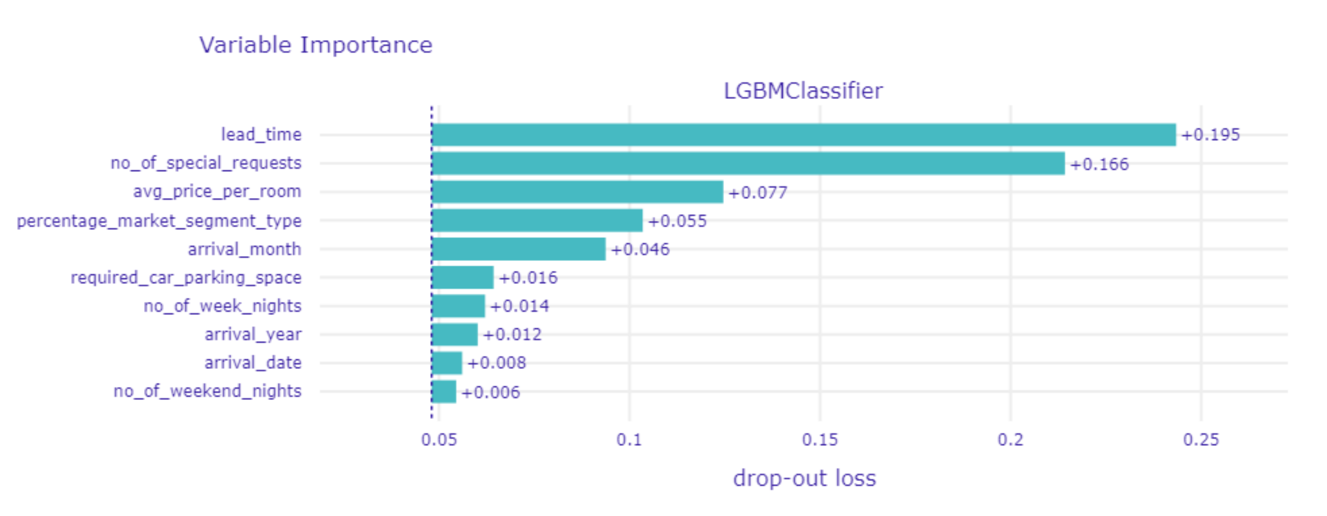
From the confusion matrix we can analyze that the model has **Actual Positive Rate (Recall)** of 80% means from 1487 customers that our model predicts customers will cancel reservation, then 80% of them or **1189 of them will exactly cancel reservation**

Learning Curve



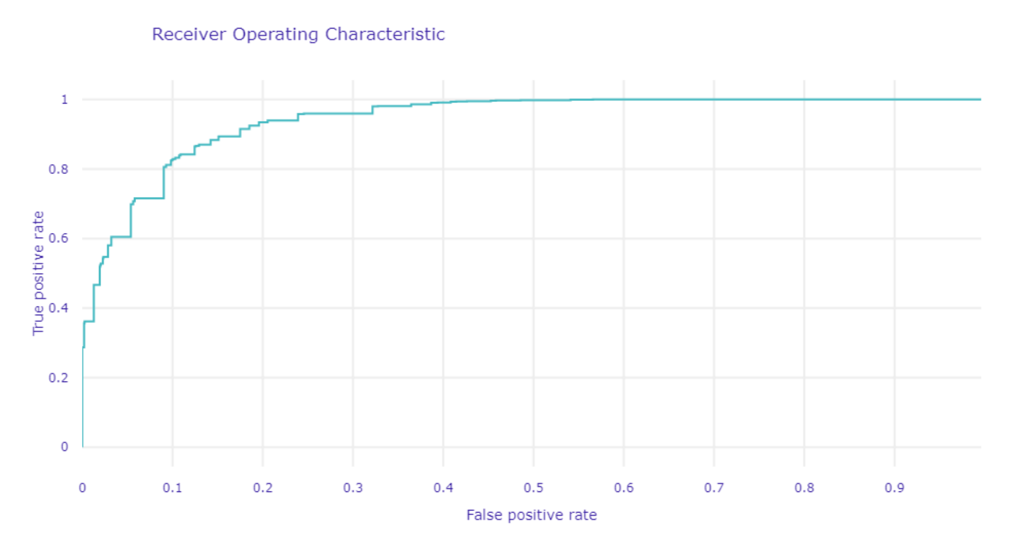
The log loss of the model is low (close to 0), it means that the model is making high accurate predictions. As the training size increases, the training score and validation score graphs become closer to each other, which indicates that the model is not overfitting.

Features Importance



Based on the calculation results, it was found that with the model, the 3 most influential feature importances are lead time, number of special requests, and average price per room

ROC Curve



The ROC curve is close to the top-left corner, indicating that the model has a good performance in discriminating the positive and negative classes

How Model Impact to Hotel Business

- Models are important for hotels as they can accurately predict the number of customers expected to attend during a given period. This information can help hotels allocate resources efficiently, such as food supplies and hotel employees, which can lead to improved operational efficiency, reduced costs, and increased customer satisfaction. By utilizing models, hotel companies can make better business decisions, which can positively impact their growth.

- Accurate predictions of the number of customers who will show up can help hotels allocate their resources more efficiently, avoiding waste and improving operational efficiency. For example, by predicting with 85% accuracy who will actually show up during a certain period, hotels can adjust their resources accordingly, thereby saving costs and increasing customer satisfaction.

Conclusion and Recommendation

5

Conclusion and Recommendation

- Monitor customers who are more likely to cancel reservations, such as new online customers without a vehicle who book far in advance. Consider offering incentives for early bookings or implementing a tiered pricing system that charges more for last-minute bookings to reduce cancellations.

- October has the highest number of reservation cancellations. The company should focus on this month and offer a special promotion, such as the Oktoberfest event, to encourage customers not to cancel. On Sundays, reservation cancellations and percentages increase, so a weekend promotion is recommended.

- Customers from the complimentary market segment have a low cancellation rate. The company can offer special promos and discounts to incentivize customers and reduce cancellations.

Conclusion and Recommendation

- Peak cancellation and reservation periods occur in October and September respectively. The company can increase rental prices during peak demand and implement a partial refund policy during the low season to minimize losses. Cancellations during peak season are not charged due to high demand.

- Sending reminders to customers about their reservations can help reduce cancellations. This can be done through various communication channels and should include all relevant reservation information.

Thank YOU

www.linkedin.com/in/muhammad-hisyam-ramadhan

ramadhan.hisyam99@gmail.com

CREDITS:

- Kaggle
- Slidesgo
- FreepikUnsplash
- Iconscout