**Project Overview**

**Title :** Predicting Hotel Reservation Cancellations

**Description :** Develop robust machine learning models to accurately predict hotel reservation cancellations. By utilizing these models, the company aims to improve revenue management, optimize resource allocation, and potentially implement targeted retention strategies.

**Benefits :**

* Revenue Optimization: By predicting potential cancellations, the hotel can

implement dynamic pricing or overbooking strategies to minimize revenue loss.

* Resource Management: Accurate cancellation predictions can help in better staff scheduling and inventory management.
* Customer Insights: Understanding factors influencing cancellations can help in improving services and customer satisfaction.

**Dataset Information**

**Dataset Details :** The dataset contains various features related to hotel bookings, including guest details, booking history, and reservation specifics. The key variables include:

* **Booking\_ID**: Unique identifier for each booking.
* **No\_of\_adults**: Number of adults in the booking.
* **No\_of\_children**: Number of children in the booking.
* **No\_of\_weekend\_nights**: Number of weekend nights booked.
* **No\_of\_week\_nights**: Number of weekday nights booked.
* **Meal\_type**: Type of meal plan booked.
* **Required\_car\_parking\_spaces**: Parking space requirement (Yes/No).
* **Room\_type\_reserved**: Type of the room reserved by the guest.
* **Lead\_time**: Days between booking and arrival.
* **Arrival\_year**: Year of arrival
* **Arrival\_month**: Monthof arrival
* **Arrival\_date**: Date of arrival
* **Market\_segment**: Source of booking (e.g., Online, Direct, Corporate).
* **Repeated\_guest**: Indicates whether the guest has booked before.
* **Previous\_cancellations**: Number of previous cancellations by the guest.
* **Previous\_booking\_not\_canceled**: Number of previous not canceled by the guest.
* **Avg\_Price\_per\_room**: Average price per day of the reservation (in euros).
* **Special\_requests**: Number of special requests made by the guest (e.g., high floor, view from the room, etc).
* **Booking\_status**: Indicates if the booking was canceled or not.

**Dataset Size** **:**

* **Rows (Data)** : 36275
* **Columns (Features)** : 19

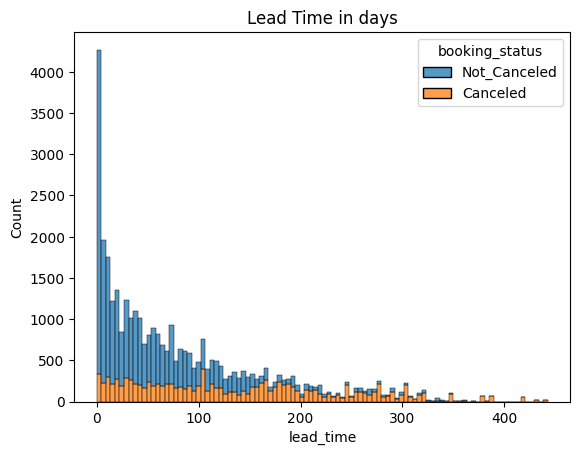
**Data Preprocessing**

**Get the basic information about the data**

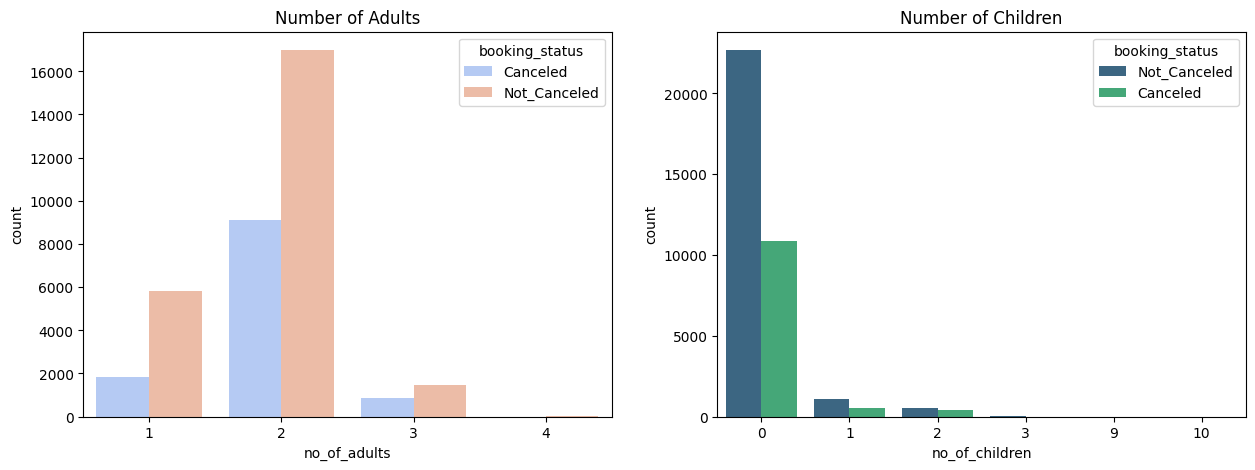
* Shape the data
* Check data type each columns
* Checking is any null value
* Checking is any duplicate value
* Checking how many category in each category columns
* Create new column date of arrival (merge three columns in one column date/month/year)
* Checking unique value in each columns
* Check statistics information

**Exploratory Data Analysis**

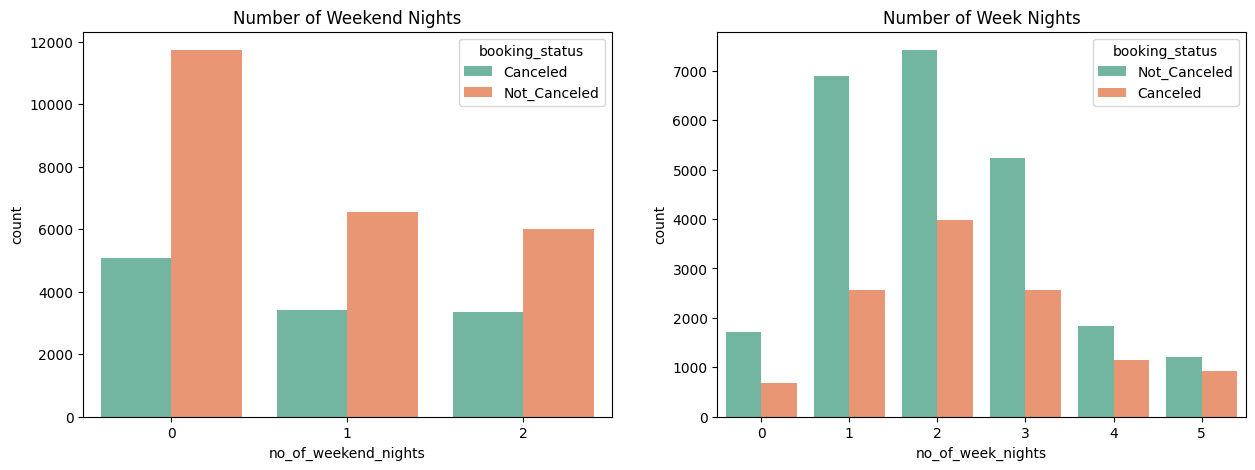
* Analyze the impact of lead time on cancellation rates.

****

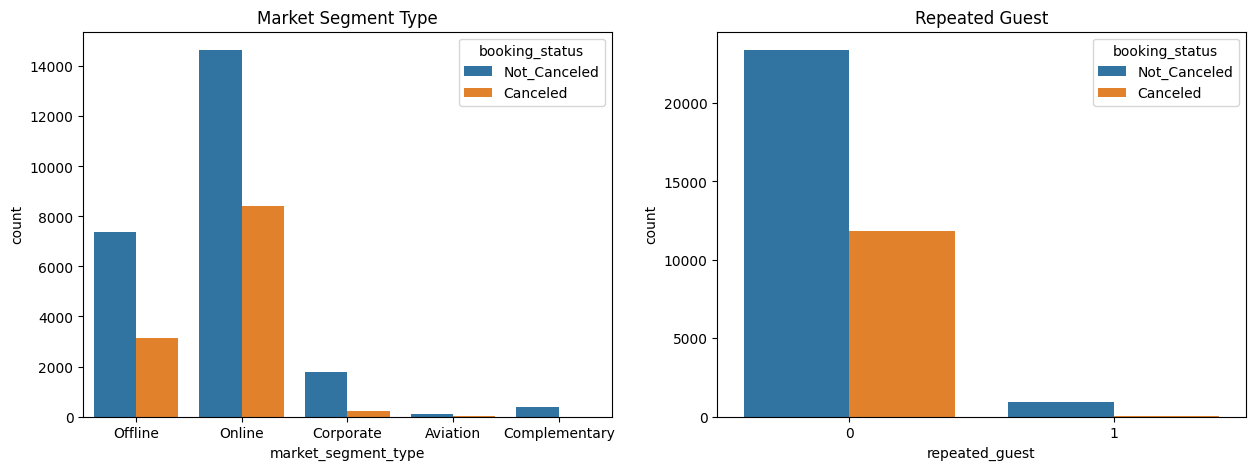
* Investigate the relationship between guest composition (adults, children) and cancellation likelihood.



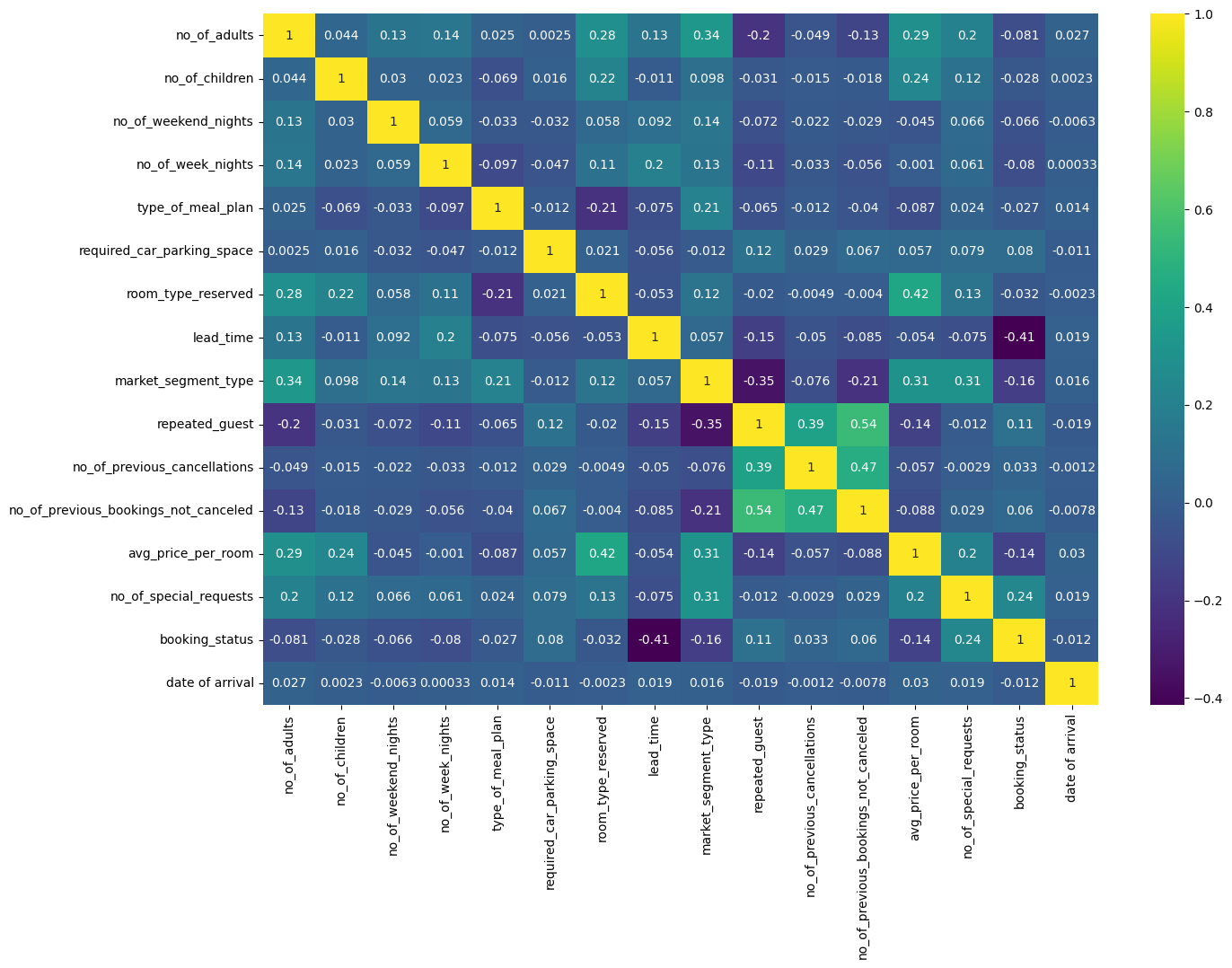
* Examine the effect of booking timing (weekdays vs. weekends) on cancellations.

****

* Study the influence of market segment and repeated guest status on cancellation patterns.

****

* Correlation Matrix



**Model Training**

**Model with their accuracy, mean absolute error and mean squared error**

1. **Logistic Regression :**

Accuracy Score: 0.80

Mean Absolute Error: 0.19

Mean Squared Error: 0.19

1. **Decision Tree :**

Accuracy Score: 0.86

Mean Absolute Error: 0.13

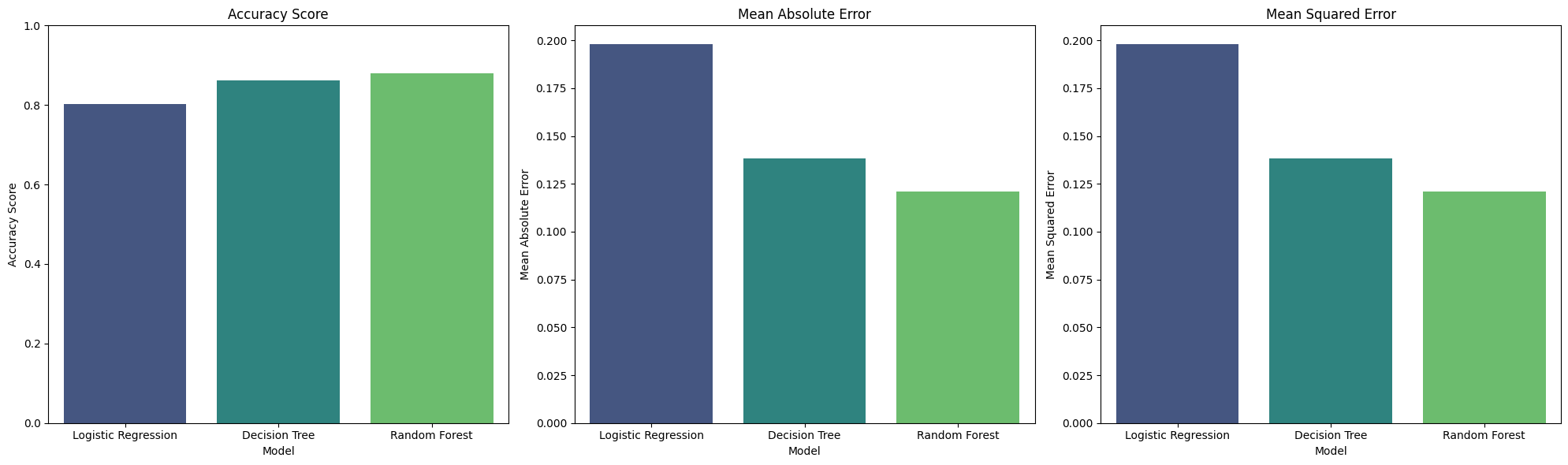
Mean Squared Error: 0.13

1. **Random Forest :**

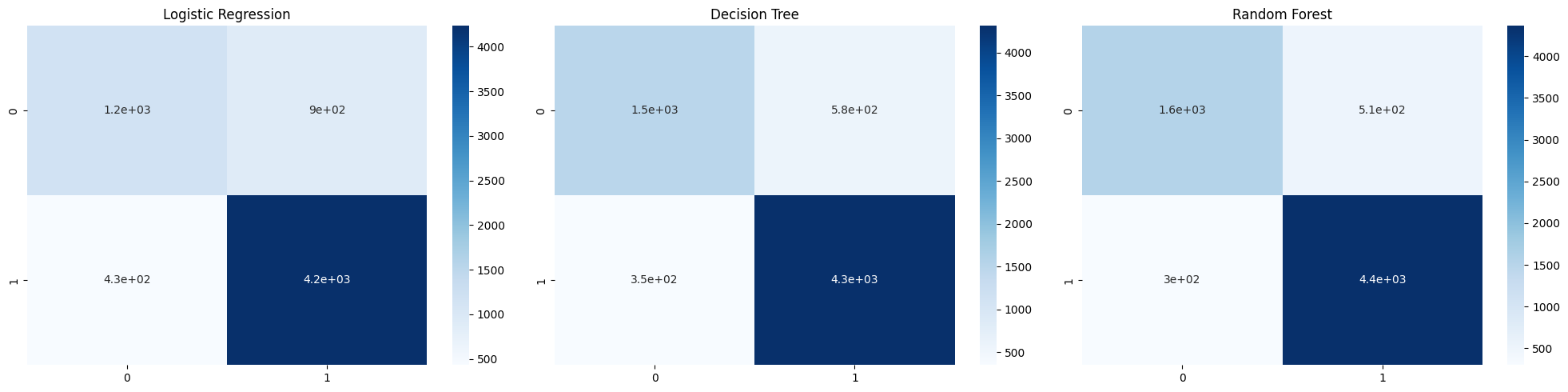
Accuracy Score: 0.87

Mean Absolute Error: 0.12

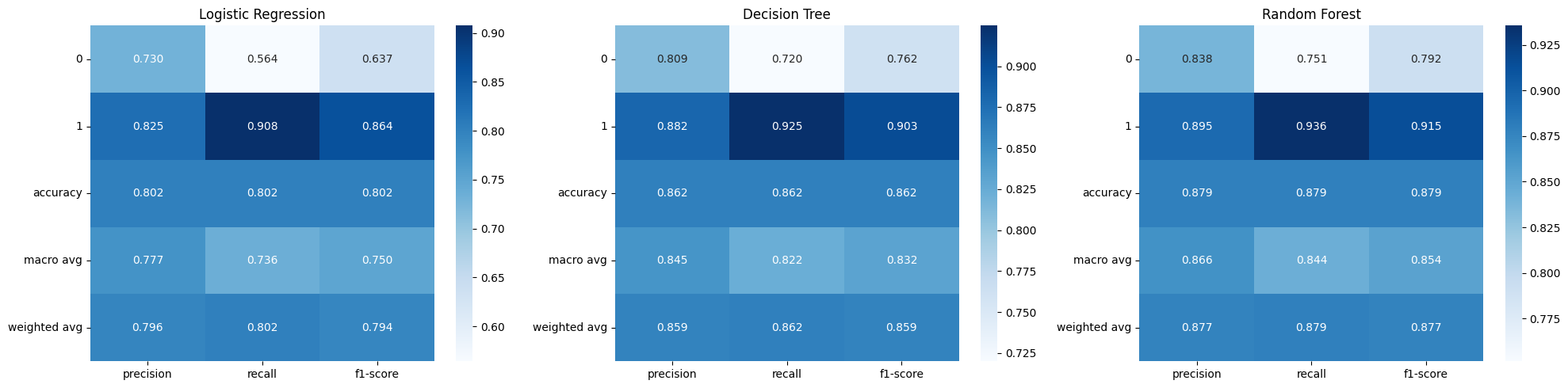
Mean Squared Error: 0.12



**Model with their confusion matrix**



**Model with their classification report**



## 

## **Conclusion**

**The analysis revealed key insights :**

* The lead time is a significant predictor of cancellations—longer lead times often result in more cancellations.
* Customers who are first-time bookers have a higher cancellation rate.
* Market segment plays a role - online bookings tend to have a higher probability of cancellation.
* The Random Forest Classifier and Decision Tree Classifier models performed best, with high accuracy and precision scores.

Predicting hotel reservation cancellations helps improve revenue management and optimize operational efficiency. The findings suggest that focusing on lead time and previous customer behavior can help mitigate cancellation risks.

## **Recommendations :**

* Implement stricter cancellation policies for long lead-time bookings.
* Offer incentives for repeated guests to reduce cancellations.
* Use predictive analytics for dynamic pricing strategies.
* Enhance data collection to include customer preferences and feedback.

This report provides valuable insights for hotel management to make data-driven decisions and enhance customer retention strategies.