HOTEL_BOOKINGS

Business Understanding

Hotel booking entails the process through which guests are able to schedule dates and length of stay, room preference among other specifications in a hotel. Bookings can be done physically at the hotel premises, through an agent or online etc. Recently, City and Resort Hotels have been affected by high cancellation rates among other booking trends. As a result, this has led to low revenue generation in those facilities. Since the goal of every business entity is to make profit, the cancellation rates should be dealt with and maintained as low as possible.

In this analysis we will address the factors which have highly contributed to high hotel cancellation rates and give recommendations to the hotel owners so that they can mitigate this risk better and earn more profit. The analysis will also be useful to the consumers of the hotel facilities and thus they will be in a position to know the ideal times of the year to book for the hotel rooms, the optimal stay length that offers them best rates etc.

The aim of this project is to analyse the Hotels data and study cancellations and their underlying patterns in order to reduce them and secure revenue.

Problem Statement

The hotel industry is a vital component of the global tourism sector, serving millions of travellers each year. However, this dynamic and competitive industry faces numerous challenges that demand innovative solutions and strategic interventions.

Hotel bookings have seen a substantial shift towards online reservations over the years, largely due to factors like the convenience offered by online platforms, the wide range of options available, and increased flexibility. Today, there are multiple methods for making hotel reservations. These include using booking agencies or directly contacting the hotels.

Our goal is to develop a predictive model capable of accurately identifying and forecasting hotel bookings and cancellations. This will be achieved through a comprehensive analysis of booking history, where we will consider factors such as: number of cancellations after reservation, number of reservations not canceled, travel agency used to make reservations, number of stays, market segmentation, whether corporate or travel agent, number of days in the waiting list, the average daily rate of the reservations, total number of special requests received, reservation status and arrival date.

This predictive model will provide valuable insights into the analysis, helping us identify the seasonal trends and factors contributing to booking cancellations, such as waiting time, to ensure a more stable revenue stream. Furthermore, we shall explore opportunities for customizing loyalty programs based on customer preferences and behaviours, ultimately increasing repeat bookings.

Main Objective

The model to be able to predict the hotel bookings and cancellations

Specific Objectives

Utilize exploratory data analysis (EDA) techniques to uncover trends and patterns in hotel bookings schedules, such as variations in checkout and arrival times. This involves a nuanced exploration of individual variables through univariate analysis, understanding relationships between pairs in bivariate analysis and unravelling intricate interactions in multivariate analysis.

Identify key variables and factors contributing to the clients' cancellations at the hotel.

Generate insights into which hotels are consistently being booked and having the most cancellations in terms of various factors such as service and quality reliability. The ultimate aim is to enhance resource allocation, optimize booking strategies and improve guest experience.

Identify factors influencing repeat bookings and develop strategies to improve customer retention.

Evaluate the performance of booking agents to identify their impact on bookings and cancellations by customers.

Data description

Hotel (object): categorical variable that indicates the type of hotel (resort or city).

Is_canceled (int64): Binary variable that indicates whether the reservation was canceled (1) or not (0).

lead_time (int64): Integer variable indicating the number of days between the date the reservation was made and the arrival date.

arrival_date_year (int64): Integer variable indicating the year of the arrival date.

arrival_date_month (object): Categorical variable indicating the month of the arrival date.

arrival_date_week_number (int64): Integer variable indicating the week number of the arrival date.

Arrival_day arrival_date_day_of_month (int64): Integer variable indicating the day of the month of the arrival date.

stays_in_weekend_nights (int64): Integer variable indicating the number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel.

stays_in_week_nights (int64): Integer variable indicating the number of week nights (Monday through Friday) the guest stayed or booked to stay at the hotel.

adults (int64): integer variable indicating the number of adults included in the reservation.

children (float64): Integer variable indicating the number of children included in the reservation.

babies (int64): integer variable that indicates the number of babies included in the reservation.

meal (object): categorical variable that indicates the type of meal reserved.

country (object): Categorical variable indicating the guest's country of origin

market_segment (object): Categorical variable indicating the guest's market segment (e.g. corporate, travel agent, etc.).

distribution_channel (object): Categorical variable indicating the distribution channel used to make the reservation.

is_repeated_guest (int64): Binary variable indicating whether the guest is a repeat guest (1) or not (0).

previous_cancellations (int64): Integer variable indicating the number of previous reservations that were canceled by the guest before the current reservation.

previous_bookings_not_canceled (int64): Integer variable indicating the number of previous reservations that were not canceled by the guest before the current reservation.

reserved_room_type (object): Categorical variable indicating the type of room reserved by the guest.

assigned_room_type (object): Categorical variable indicating the type of room assigned to the guest at check-in.

booking_changes (int64): Integer variable indicating the number of changes/modifications made to the reservation from the time the reservation was made until the guest arrived at the hotel.

deposit_type (object): Categorical variable indicating the type of deposit made to guarantee the reservation.

agent (float64): Integer variable indicating the ID of the travel agent who made the reservation.

company (float64): integer variable indicating the ID of the company or entity that made the reservation.

days_in_waiting_list (int64): Integer variable indicating the number of days the reservation was on the waiting list before being confirmed to the guest.

customer_type (object): Categorical variable indicating the type of reservation (for example, transient, contract, group, etc.).

adr (float64): Float variable indicating the average daily rate (i.e. the total reservation amount divided by the number of nights stayed) of the reservation.

required_car_parking_spaces (int64) – Integer variable indicating the number of parking spaces required by the guest.

total_of_special_requests (int64): Integer variable indicating the number of special requests made by the guest during the reservation or during their stay at the hotel. For example, a special request could be a room with a king-size bed or a request for additional room service.

reservation_status (object): Categorical variable that indicates the latest status of the reservation, such as canceled, check-in, check-out, etc.

reservation_status_date (object): Date variable indicating the date the reservation's last status was set.

arrival_date (object): Date variable that indicates the guest's arrival date at the hotel.

year_month (object): period variable that indicates the year and month of the guest's arrival date at the hotel.

Name (object): Categorical variable indicating the name of the guest

Email(object): Categorical variable indicating the email address of the guest

Phone_number

Credit_card

Metric of Success

For our model we will be using the F1 score, Recall, Precision and accuracy.

Data sources

Our data is from Kaggle

Link of the dataset

https://www.kaggle.com/datasets/sraddhanjalibarik/hotel-booking-data-set

Research Questions

What are the variables that affect hotel reservation cancellation?

Where are the majority booking? online or offline?

How can we assist the hotel in making their pricing and promotional decisions?

Why are most reservations made by agents offline?

Which distribution channels and market segments show the highest booking success rate?

How can booking strategies be optimized to minimize cancellations and maximize revenue?

What personalized services can be introduced to enhance guest experience?

What strategies can be implemented to improve collaboration with high-performing partners?

What are the key preferences and behaviours of repeated guests?