

Hotel Bookings Management Report

Problem statement

Hotel booking cancellations pose significant challenges for both customers and hotel management. The ability to predict these cancellations accurately can revolutionize resource planning, revenue management, and customer satisfaction within the hospitality industry. Efficiently predicting and managing hotel room occupancy rates is crucial for maximizing profits while ensuring customer satisfaction. Understanding the factors influencing booking patterns and guest preferences is essential for hoteliers to make informed decisions regarding pricing, promotions, and inventory management.

Motivation

Competitive Advantage: Understanding market trends and customer preferences can provide a competitive edge by offering tailored services and pricing.

Improve Operations: Predictive analytics can help hotels to manage inventory more effectively, preventing overbooking or underutilization of rooms.

Revenue Management: Accurate predictions help in maximizing revenue by allowing hotels to resell canceled rooms or offer them to waitlisted customers, preventing revenue loss.

Optimizing Resource Utilization: Anticipating cancellations enables hotels to optimize resource allocation, such as staff scheduling, room inventory management, and service provision, reducing operational costs.

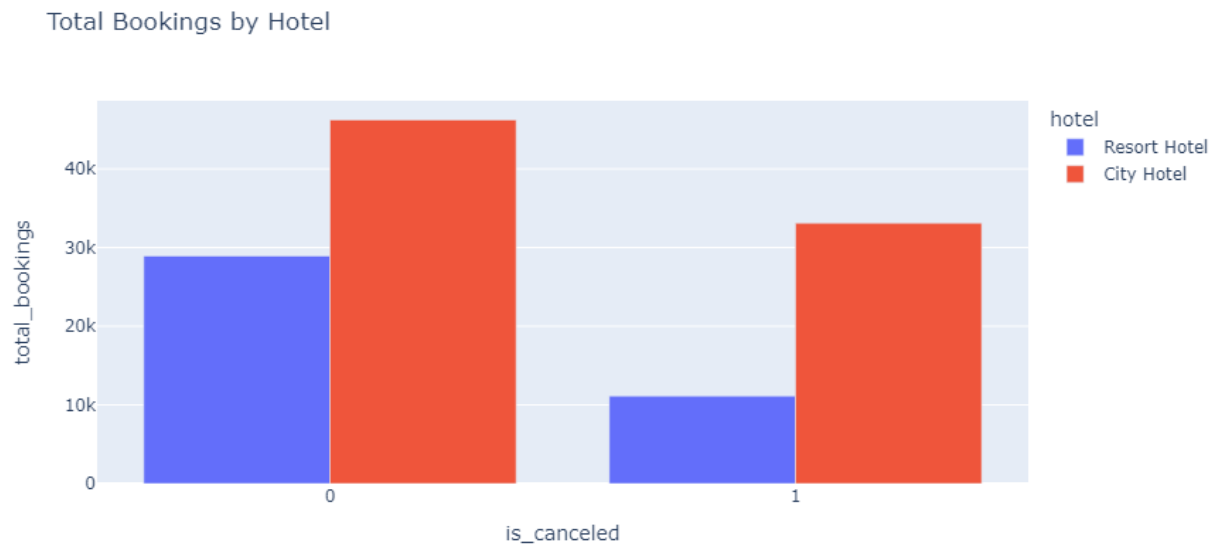
Data-Driven Insights: Hotel booking data, including historical bookings, customer profiles, booking channels, and reservation timing, contains valuable insights. Utilizing this data to predict cancellations can improve decision-making.

Data preprocessing

A dataset of approximately 110000 records with 36 attributes over the years 2015-2017 have been selected for our analysis which consists of various information about the customer and booking preferences. In which we have some attributes with null values but has been taken care of during normalization.

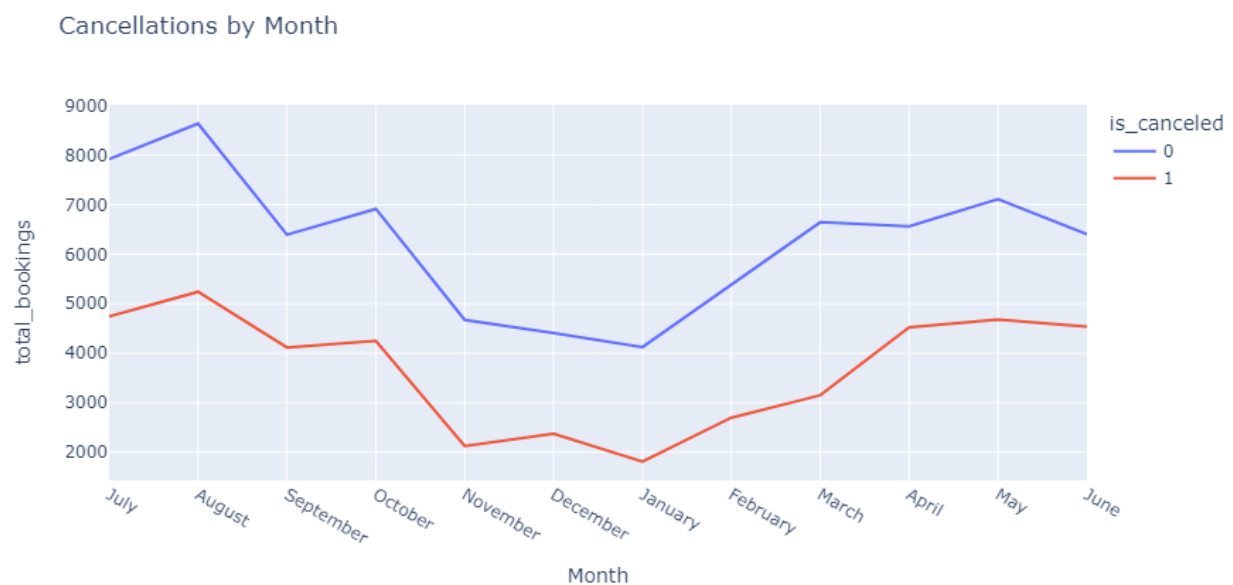
We have created 3 tables namely Hotel, Customer, Bookings

The main attribute of booking status is **'is_canceled'** with 0 or 1 indicates if the bookings is not canceled or cancelled respectively.



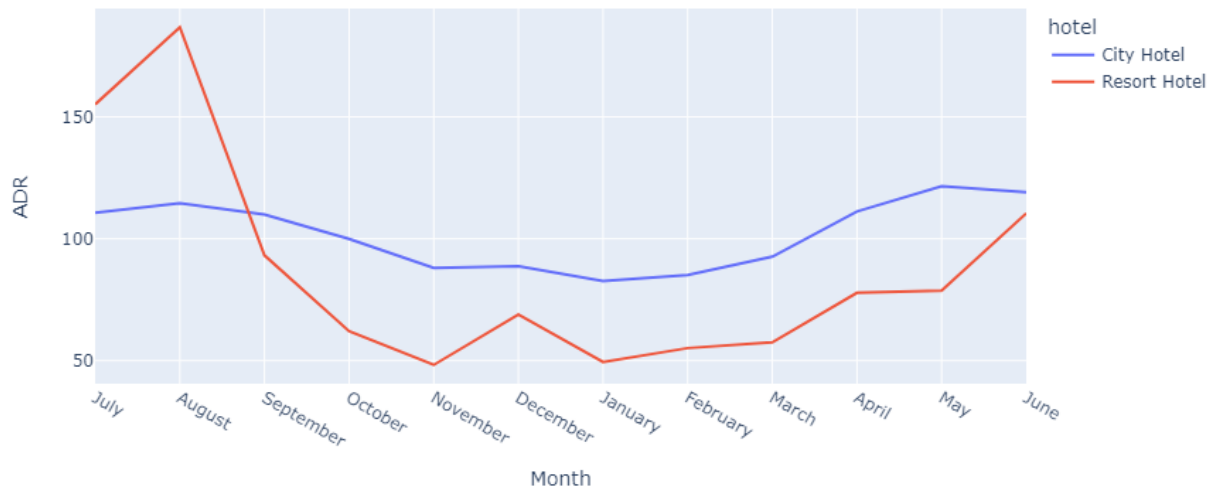
Above graphs depict non-canceled bookings of Resort Hotel are 72 % of total bookings and canceled bookings are of 28%. Whereas 58 % of total bookings for City Hotel are not cancelled while 42% were cancelled.

After analyzing the average cancellations per month, it is observed that in August most bookings and cancellations were happened and January with the least bookings and cancellations.



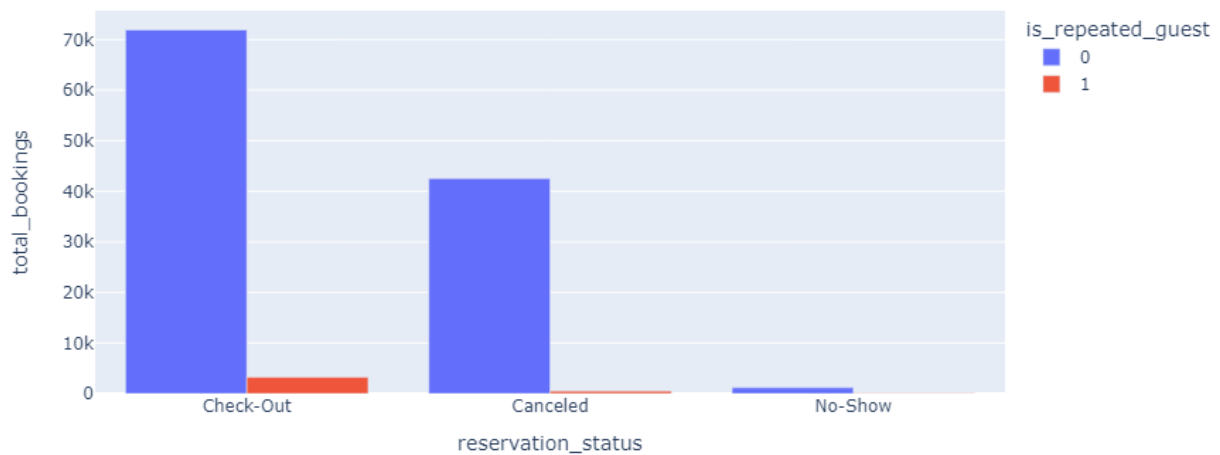
The reason for august to have the most cancellations and January to have least is because of their Average Daily Rates (ADR). Looking at the ADR per month gives us the moving average of ADR each month.

Average Daily Rates by Month

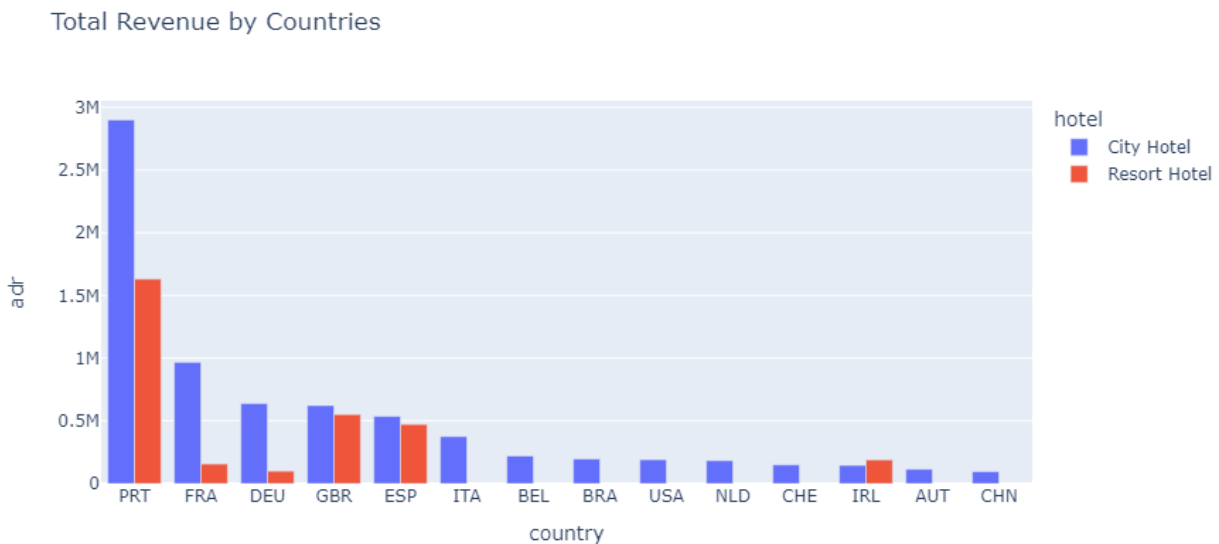


The total bookings are grouped by reservation status and whether the guest is repeated. Non-repeated guest's check-out rates and cancelations are very high compared to the repeated guests.

Reservation status of Repeated Guests

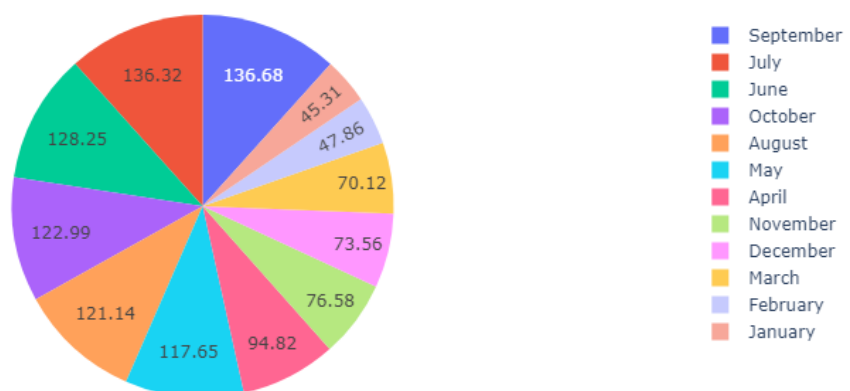


Below graphs explains the total revenue by each country over the years and it is evident travelers from **Portugal** are responsible for most the revenue in both hotels. Also, people from France and Germany mostly tend to stay in City hotels where as people from London and Spain are similarly distributed in both hotels.



Finally, we have lead time which is the elapsed days between the booking date and arrival date. We have identified that **July** and **September** are the busy months as people have to book at least **136** days prior to their arrival date and January to have the least lead time of 45 days.

Elapsed days between booking date and arrival date



Above are the exploratory data analysis we performed to identify the main attributes which may be responsible for hotel bookings and their cancellations. Now in order for hotels to improve revenue for

their better services and hospitality, we came up few classifications models to predict the attributes responsible for booking cancellations.

lead_time	0.260322
adr	0.200390
arrival_date_week_number	0.120376
stays_in_weekend_nights	0.118189
total_of_special_requests	0.078303
stays_in_week_nights	0.061112
previous_cancellations	0.057056
booking_changes	0.037811
required_car_parking_spaces	0.027697
adults	0.021218
days_in_waiting_list	0.009009
previous_bookings_not_canceled	0.008517

So, **lead_time** and **ADR** are crucial factors while predicting a booking might be cancelled or not. And, similarly remaining feature's importance been be shown in the order.

Also, **Random Forest classifier** has the highest accuracy of **84.49%** and precision of 92% along with 88% of F1-score. In our comparison, Random Forest outperformed other 3 models in terms of accuracy and precision. So, it is suggested to utilize Random Forest for their hotel bookings classifications.

Comparison of various models

