

Context

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate?

What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, and the number of adults, children, and the number of available parking spaces, among other things.

Features

variable	class	description		
hotel	character	Hotel (H1 = Resort Hotel or H2 = City Hotel)		
is_canceled	double	Value indicating if the booking was canceled (1) or not (0)		
lead_time	double	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date		
arrival_date_year double		Year of arrival date		
arrival_date_month	character	Month of arrival date		
arrival_date_week_number	double	Week number of year for arrival date		
arrival_date_day_of_month double		Day of arrival date		
stays_in_weekend_nights double		Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel		
stays_in_week_nights	double	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel		
adults	double	Number of adults		
children	double	Number of children		

babies	double	Number of babies	
meal	character	Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)	
country	character	Country of origin. Categories are represented in the ISO 3155–3:2013 format	
market_segment	character	Market segment designation. In categories, the term "TA" means "Travel Agents" and "T means "Tour Operators"	
distribution_channel	character	Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"	
is_repeated_guest	double	Value indicating if the booking name was from a repeated guest (1) or not (0)	

previous_cancellations	double	Number of previous bookings that were cancelled by the customer prior to the current booking		
previous_bookings_not_canceled	double	Number of previous bookings not cancelled by the customer prior to the current booking		
reserved_room_type	character	Code of room type reserved. Code is presented instead of designation for anonymity reasons		
assigned_room_type	character	Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) by customer request. Code is presented instead of designation for anonymity reasons		
booking_changes	double	Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation		
deposit_type characte		Indication on if the customer made a deposit to guarantee the booking. This variable c assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.		
agent	character	ID of the travel agency that made the booking		
company characte		ID of the company/entity that made the booking or responsible for paying the booking ID is presented instead of designation for anonymity reasons		
days_in_waiting_list	double	Number of days the booking was in the waiting list before it was confirmed to the customer		
customer_type character		Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it. Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking		
adr	double	Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights		
required_car_parking_spaces	double	Number of car parking spaces required by the customer		
total_of_special_requests	double	Number of special requests made by the customer (e.g. twin bed or high floor)		

HOTEL BOOKING DEMAND ANALYSIS

Project Proposal

reservation_status	character	Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – customer did not check-in and did inform the hotel of the reason why		
reservation_status_date	double	Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel		

Data Head

Hotel = pd.read_csv("hotel_bookings.csv") Hotel.head()								
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	
	Resort Hotel		342	2015	July			
	Resort Hotel		737	2015	July			
2	Resort Hotel			2015	July			
3	Resort Hotel		13	2015	July			
4	Resort Hotel		14	2015	July			
	ows × 32	columns						

Data Dimension

```
In [6]: Hotel.shape
(119390, 32)
```

The Dataset has 119390 rows and 32 columns (13 Categorical and 19 Numerical)

Data Info

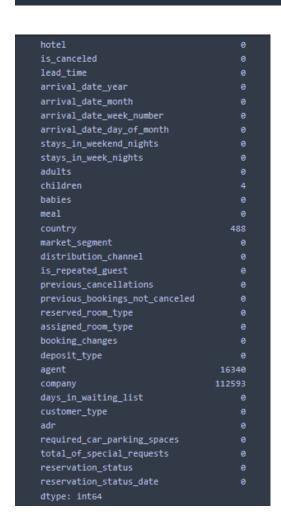
```
23 agent 103050 non-null float64
24 company 6797 non-null float64
25 days_in_waiting_list 119390 non-null int64
26 customer_type 119390 non-null object
27 adr 119390 non-null float64
28 required_car_parking_spaces 119390 non-null int64
29 total_of_special_requests 119390 non-null int64
30 reservation_status 119390 non-null object
31 reservation_status_date 119390 non-null object
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

HOTEL BOOKING DEMAND ANALYSIS

Project Proposal

Now, let's extract some information like the number of non-null values in every column.

```
np.sum(Hotel.isnull())
```



Problems in Data

- Object Data Type of "reservation_status_date" must be converted to DateTime.
- The Dataset contains 4 columns with null values.