

Hotel Booking Demands



Agenda

- About our dataset
- Dataset attributes
- Dataset characteristics
- Analysis questions



About our dataset



This data set contains booking information for a city hotel and a resort hotel.



It includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.



The data is originally from the article [Hotel Booking Demand Datasets](#).



It consists of 119390 rows and 32 columns.

Data attributes

- **Hotel:** represents whether it is a resort or city hotel.
- **Is_canceled:** indicated if the booking was cancelled (1) or not (0).
- **lead_time :** Number of days that elapsed between the entering date of the booking and the arrival date
- **arrival_date_year :** Year of arrival date
- **arrival_date_month :** month of arrival date
- **arrival_date_week_number:** Week number of year for arrival date
- **arrival_date_day_of_month:** Day of arrival date
- **stays_in_weekend_nights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

Data attributes cont.

- **stays_in_week_nights:** Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **Adults:** Number of adults
- **Children:** Number of children
- **Babies:** Number of babies
- **Meal:** Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal
- **Country:** Country of origin. Categories are represented in the ISO 3155–3:2013 format
- **market segment:** Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **distribution_channel:** Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **stays_in_week_nights:** Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **is_repeated_guest:** Value indicating if the booking name was from a repeated guest (1) or not (0)

Data attributes cont.

- **previous_cancellations:** Number of previous bookings that were cancelled by the customer prior to the current booking
- **previous_bookings_not_canceled:** Number of previous bookings not cancelled by the customer prior to the current booking
- **reserved_room_type:** Code of room type reserved. Code is presented instead of designation for anonymity reasons.
- **assigned_room_type:** Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due
- **booking_changes:** Number of changes/amendments made to the booking from the moment the booking was entered on the PMS
- **deposit_type:** Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was
- **Agent:** ID of the travel agency that made the booking
- **Company:** ID of the company/entity that made the booking or responsible for paying the booking.

Data attributes cont.

- **days_in_waiting_list:** Number of days the booking was in the waiting list before it was confirmed to the customer
- **customer_type:** Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of
- **Adr:** Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
- **required_car_parking_spaces:** Number of car parking spaces required by the customer
- **total_of_special_requests:** Number of special requests made by the customer (e.g. twin bed or high floor)
- **reservation_status:** Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out
- **reservation_status_date:** Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to

Dataset characteristics

Our dataset contains:

1. **Missing values:** resides in columns 'country' , 'agent' and 'company'.
2. **Categorical data:** in multiple columns such as reservation_status and customer_type
3. **Outliers:** multiple columns contains outliers such as 'adr' and 'total_of_special_requests'
4. **Imbalanced:** our target column 'is_canceled' is imbalanced.

Aim of the project:

- The target is to predict whether the client will cancel his reservation or not.

Analysis Questions

1. Which month had the most reservations?
2. How many clients have cancelled their booking?
3. What is the average number of children per booking?
4. How many loyal clients do we have?
5. Which travel agency have made the most bookings?
6. Which room type is reserved the most?