**HOTEL BOOKINGS ANALYSIS DOCUMENTATION**

This data was downloaded from [Kaggle](http://kaggle.com/datasets/jessemostipak/hotel-booking-demand).

**Analysis Process**

1. Create a working copy in Excel for easier profiling

* A quick profiling revealed the following about the data columns
  1. Hotel – there are two hotels with no blanks (Resort Hotel and City Hotel)
  2. Canceled: 0 & 1

Assumptions

* + 0 – No
  + 1 – Yes
  1. Lead Time: integer column, no blanks.
  + Assumption: lead time is in minutes
  1. Arrival date year - integer column, no blanks, representing year
  2. Arrival date month – text column, no blanks, representing month
  3. Arrival date week number – integer, no blanks, representing week number
  4. Arrival date day of month – integer, no blanks, representing actual day.

Below is the data definition from the referenced original download:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type** | **Description** | **Source/Engineering** |
| ***ADR*** | Numeric | [Average Daily Rate as defined by [5]](https://www.sciencedirect.com/science/article/pii/S2352340918315191" \l "bib5) | BO, BL and TR / Calculated by dividing the sum of all lodging transactions by the total number of staying nights |
| ***Adults*** | Integer | Number of adults | BO and BL |
| ***Agent*** | Categorical | [ID of the travel agency that made the bookinga](https://www.sciencedirect.com/science/article/pii/S2352340918315191" \l "tbl1fna) | BO and BL |
| ***ArrivalDateDayOfMonth*** | Integer | Day of the month of the arrival date | BO and BL |
| ***ArrivalDateMonth*** | Categorical | Month of arrival date with 12 categories: “January” to “December” | BO and BL |
| ***ArrivalDateWeekNumber*** | Integer | Week number of the arrival date | BO and BL |
| ***ArrivalDateYear*** | Integer | Year of arrival date | BO and BL |
| ***AssignedRoomType*** | Categorical | 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) or by customer request. Code is presented instead of designation for anonymity reasons | BO and BL |
| ***Babies*** | Integer | Number of babies | BO and BL |
| ***BookingChanges*** | Integer | 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 | BO and BL/Calculated by adding the number of unique iterations that change some of the booking attributes, namely: persons, arrival date, nights, reserved room type or meal |
| ***Children*** | Integer | Number of children | BO and BL/Sum of both payable and non-payable children |
| ***Company*** | Categorical | ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons | BO and BL. |
| ***Country*** | Categorical | [Country of origin. Categories are represented in the ISO 3155–3:2013 format [6]](https://www.sciencedirect.com/science/article/pii/S2352340918315191" \l "bib6) | BO, BL and NT |
|  |  |  |  |
| ***CustomerType*** | Categorical | Type of booking, assuming one of four categories: | BO and BL |
|  |  | 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 |  |
| ***Days in Waiting List*** | Integer | Number of days the booking was in the waiting list before it was confirmed to the customer | BO/Calculated by subtracting the date the booking was confirmed to the customer from the date the booking entered on the PMS |
|  |  |  |  |
| ***DepositType*** | Categorical | Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: | BO and TR/Value calculated based on the payments identified for the booking in the transaction (TR) table before the booking׳s arrival or cancellation date. |
|  |  | No Deposit – no deposit was made; |  |
|  |  |  | In case no payments were found the value is “No Deposit”. |
|  |  |  | If the payment was equal or exceeded the total cost of stay, the value is set as “Non Refund”. |
|  |  | Non Refund – a deposit was made in the value of the total stay cost; |  |
|  |  |  | Otherwise the value is set as “Refundable” |
|  |  | Refundable – a deposit was made with a value under the total cost of stay. |  |
| ***DistributionChannel*** | Categorical | Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators” | BO, BL and DC |
| ***IsCanceled*** | Categorical | Value indicating if the booking was canceled (1) or not (0) | BO |
| ***IsRepeatedGuest*** | Categorical | Value indicating if the booking name was from a repeated guest (1) or not (0) | BO, BL and C/ Variable created by verifying if a profile was associated with the booking customer. If so, and if the customer profile creation date was prior to the creation date for the booking on the PMS database it was assumed the booking was from a repeated guest |
| ***LeadTime*** | Integer | Number of days that elapsed between the entering date of the booking into the PMS and the arrival date | BO and BL/ Subtraction of the entering date from the arrival date |
| ***MarketSegment*** | Categorical | Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators” | BO, BL and MS |
|  |  |  |  |
| ***Meal*** | Categorical | Type of meal booked. Categories are presented in standard hospitality meal packages: | BO, BL and ML |
|  |  | 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) |  |
| ***PreviousBookingsNotCanceled*** | Integer | Number of previous bookings not cancelled by the customer prior to the current booking | BO and BL / In case there was no customer profile associated with the booking, the value is set to 0. Otherwise, the value is the number of bookings with the same customer profile created before the current booking and not canceled. |
| ***PreviousCancellations*** | Integer | Number of previous bookings that were cancelled by the customer prior to the current booking | BO and BL/ In case there was no customer profile associated with the booking, the value is set to 0. Otherwise, the value is the number of bookings with the same customer profile created before the current booking and canceled. |
| ***RequiredCardParkingSpaces*** | Integer | Number of car parking spaces required by the customer | BO and BL |
|  |  |  |  |
| ***ReservationStatus*** | Categorical | Reservation last status, assuming one of three categories: | BO |
|  |  | 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 |  |
| ***ReservationStatusDate*** | Date | 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 | BO |
| ***ReservedRoomType*** | Categorical | Code of room type reserved. Code is presented instead of designation for anonymity reasons | BO and BL |
| ***StaysInWeekendNights*** | Integer | Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel | BO and BL/ Calculated by counting the number of weekend nights from the total number of nights |
| ***StaysInWeekNights*** | Integer | Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel | BO and BL/Calculated by counting the number of week nights from the total number of nights |
| ***TotalOfSpecialRequests*** | Integer | Number of special requests made by the customer (e.g. twin bed or high floor) | BO and BL/Sum of all special requests |

1. Objectives

* Understand bookings per hotel distribution.
  + Total Bookings
  + Average Bookings
  + Previous cancelled/ non-cancelled bookings
    - Sum cancellations
    - Count cancellations
  + Previous Bookings not Cancelled
    - Sum cancellations
    - Count cancellations
  + Market segment booking distribution
  + Adults, Children, Babies distribution.
    - Sum of each
* Hotel Booking trend throughout the periods: years, quarter & months. To give insights into busy & low times of the year.
* Bookings by:
  + Hotel (categories)
    - Count
  + Country (Categories)
    - Count by channel
  + Distribution channel (categories)
    - Count by channel
  + Repeat guest: Yes/ No
    - Count
  + Reserved room type
    - Count by type
  + Assigned room type
    - The count no of times reserved was same as assigned vs. the count of times it was different.
    - Add column to compare the two columns
  + Booking Changes
    - Sum
  + Deposit type (categories)
    - Count
    - Average
  + Market Segment (categories)
    - Count by category
  + Agent (categories)
    - Count
  + Company (categories)
    - Count
  + Customer Type (categories)
    - Count
  + Required parking access (Integer)
    - Sum
  + Number of special requests (Integer)
    - Sum
  + Reservation status (categories)
    - Count
  + Meal
    - Count of each meal type

1. **Transformations**
2. Date Table
   1. Create a Date table, by extracting distinct combinations of day, month, and year.
   2. Marked the table as a Date Table
   3. Created a Hierarchy
   4. Removed summation on Day & Year fields from the Model View.
   5. Added date column on fact table, created relationship of One-to-Many between Dates and Fact tables
3. Created Measures Table to house all measures

* Total Bookings: counts all bookings, using hotel column