**Nermien Elassy**

**Description of the data**

**Who:** Observations: Two Hotels (A city hotel and a resort hotel).

**When:** The data was collected in February 2019.

**Where:**

This data was originally written in February 2019 <https://www.sciencedirect.com/science/article/pii/S2352340918315191> by Nuno Antonio, Ana Almeida, and Luis Nunes. Then it was cleaned and sorted in a table by Thomas Mock and Antoine Bichat on 11th of February 2020 <https://www.kaggle.com/jessemostipak/hotel-booking-demand>.

**Why:** This data was collected due to its important role in different fields like data mining, machine learning and revenue management for educational and research purposes. Moreover, the real business data that are available for scientific researchers are rare.

**What?** Here are the 32 variables in my data with their description.

Important note: The whole data set is only analyzed from the perspective of the type of the variables and their unit of measurement. The coding part and the conclusion is only focused on specific variables that matter for my analysis to answer the two specific questions I mentioned in my report. Moreover, there are other variables that are included in my coding part for just explanation of some specific commands (Numerical summaries).

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type** | **Unit of measurement** | **Description** |
| Hotel | Categorical (binomial) (city, resort) | hotel | Hotel (H1 = Resort Hotel or H2 = City Hotel) |
| is \_canceled | Logical | 0 = false, 1=true |  |
| lead\_time | Numeric(Quantitative) | days | Number of days that elapsed between the entering date of the booking into the PMS and the arrival date |
| arrival\_date\_month | Categorical (multinomial) (January-December) | month | Month of arrival date |
| arrival\_date\_week\_number | Numeric(Quantitative) | Number of the week | Week number of year for arrival date |
| arrival\_date\_day\_of\_month | Numeric(Quantitative) | Number of the dayday | Day of arrival date |
| stays\_in\_weekend\_nights | Numeric(Quantitative) | Number of Weekend nights | Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel |
| stays\_in\_week\_nights | Numeric(Quantitative) | Number of Week nights | Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel |
| adults | Numeric(Quantitative) | Number of adults |  |
| distribution\_channel | Multinomial (TA/TO/direct/other) | channel | Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators” |
| is\_repeated\_guest | Logical | (0 = false, 1=true) |  |
| previous\_cancellations | Numeric(Quantitative) | Number of previous cancellations | Number of previous bookings that were cancelled by the customer prior to the current booking |
| reserved\_room\_type | Character string | No unit of measurement | Code of room type reserved. Code is presented instead of designation for anonymity reasons. |
| assigned\_room\_type | Character string | No unit of measurement | Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type. |
| reservation\_status | Categorical(Check-out, Cancelled, other) | Reservation status | Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer. |
| reservation\_status\_date | Character string | No unit of measurement | Date at which the last status was set. This variable can be used in conjunction with the Reservation Status |
| total\_of\_special\_requests | Numeric(Quantitative) | Number of special requests | Number of special requests made by the customer (e.g. twin bed or high floor) |
| required\_car\_parking\_spaces | Numeric(Quantitative) | Number of car parking spaces | Number of car parking spaces required by the customer |
| adr | Numeric(Quantitative) | No unit of measurement | Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights |
| customer\_type | Categorical(multinomial) (transient, transient-party, other) | Customer | Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; |
| days\_in\_waiting\_list | Numeric(Quantitative) | Number of days | Number of days the booking was in the waiting list before it was confirmed to the customer |
| company | Numeric(quantitative) | ID | ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for |
| agent | Numeric(quantitative) | ID | ID of the travel agency that made the booking |
| booking\_changes | Numeric(quantitative) | Number of changes/amendments | Number of changes/amendments made to the booking from the moment the booking was entered on the PMS |
| deposit\_type | Categorical(binomial) | No unit of measurement | Indication on if the customer made a deposit to guarantee the booking. |
| previous\_bookings\_not\_canceled | Numeric(quantitative) | Number of bookings not canceled | Number of previous bookings not cancelled by the customer prior to the current booking |
| Children | Numeric(Quantitative) | Number of children |  |
| babies | Numeric(Quantitative) | Number of babies |  |
| arrival\_date\_year | Numeric(Quantitative) | Year number | Year of arrival date |
| meal | Categorical(multinomial (BB,HB,other) | meal | Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal |
| Countries | Categorical(mutlinomial(PRT,GBR,Other) | country | Country of origin. Categories are represented in the ISO 3155–3:2013 format |
| Market\_segment | Categorical(multinomial)(online, offline, other) | Market segment | Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators |