A hotel's customers dataset

Context

This real-world customer dataset with 31 variables describes 83,590 instances (customers) from a hotel in Lisbon, Portugal.

Content

The data comprehends three full years of customer personal, behavioral, demographic, and geographical information.

Acknowledgements

Additional information on this dataset can be found in the article A Hotel's customers personal, behavioral, demographic, and geographic dataset from Lisbon, Portugal (2015-2018), written by Nuno Antonio, Ana de Almeida, and Luis Nunes for Data in Brief (online November 2020).

Data Dictionary:

Variable	Type	Description
ID	Numeric	Customer ID
Nationality	Categorical	Country of origin. Categories are represented in the ISO 3155–3:2013 format
Age	Numeric	Customer's age (in years) at the last day of the extraction period.
DaysSinceCreation	Number	Number of days since the customer record was created (number of days elapsed between the creation date and the last day of the extraction
NameHash	Categorical	Name of the customer's SHA2–256 hash string. A hash-string is the string resulting from a mathematical function that maps a string of arbitrary length to fixed-length [2]. Hash functions are used for different purposes. In this case, to allow customer's anonymization.
DocIDHash	Categorical	SHA2–256 hash-string of the identification document number the customer provided at check-in (passport number, national ID card number, or other)
AverageLeadTime	Numeric	The average number of days elapsed between the customer's booking date and arrival date. In other words, this variable is calculated by dividing the sum of the number of days elapsed between the moment each booking was made and its arrival date, by the total of bookings made by the customer
LodgingRevenue	Numeric	Total amount spent on lodging expenses by the customer (in Euros). This value includes room, crib, and other related lodging expenses
OtherRevenue	Numeric	Total amount spent on other expenses by the customer (in Euros). This value includes food, beverage, spa, and other expenses
BookingsCanceled	Numeric	Number of bookings the customer made but subsequently canceled (the costumer informed the hotel he/she would not come to stay)
BookingsNoShowed	Numeric	Number of bookings the customer made but subsequently made a "no-show" (did not cancel, but did not check-in to stay at the hotel)
BookingsCheckedIn	Numeric	Number of bookings the customer made, and which end up with a staying
PersonsNights	Numeric	The total number of persons/nights that the costumer stayed at the hotel. This value is calculated by summing all customers checked-in bookings' persons/nights. Person/nights of each booking is the result of the multiplication of the number of staying nights by the sum of adults and children

RoomNights	Numeric	Total of room/nights the customer stayed at the hotel (checked-in bookings). Room/nights are the multiplication of the number of rooms of each booking by the number of nights of the booking
DaysSinceLastStay	Numeric	The number of days elapsed between the last day of the extraction and the customer's last arrival date (of a checked-in booking). A value of 1 indicates the customer never stayed at the hotel
DaysSinceFirstStay	Numeric	the customer's first arrival date (of a checked-in booking). A value ofThe number of days elapsed between the last day of the extraction and 1 indicates the customer never stayed at the hotel
DistributionChannel	Categorical	Distribution channel usually used by the customer to make bookings at the hotel
MarketSegment	Categorical	Current market segment of the customer
SRHighFloor	Boolean	Indication if the customer usually asks for a room on a higher floor (0: No, 1: Yes)
SRLowFloor	Boolean	Indication if the customer usually asks for a room on a lower floor (0: No, 1: Yes)
SRAccessibleRoom	Boolean	Indication if the customer usually asks for an accessible room (0: No, 1: Yes)
SRMediumFloor	Boolean	Indication if the customer usually asks for a room on a middle floor (0: No, 1: Yes)
SRBathtub	Boolean	Indication if the customer usually asks for a room with a bathtub (0: No, 1: Yes)
SRShower	Boolean	Indication if the customer usually asks for a room with a shower (0: No, 1: Yes)
SRCrib	Boolean	Indication if the customer usually asks for a crib (0: No, 1: Yes)
SRKingSizeBed	Boolean	Indication if the customer usually asks for a room with a king-size bed (0: No, 1: Yes)
SRTwinBed	Boolean	Indication if the customer usually asks for a room with a twin bed (0: No, 1: Yes)
SRNearElevator	Boolean	Indication if the customer usually asks for a room near the elevator (0: No, 1: Yes)
SRAwayFromElevator	Boolean	Indication if the customer usually asks for a room away from the elevator (0: No, 1: Yes)
SRNoAlcoholInMiniBar	Boolean	Indication if the customer usually asks for a room with no alcohol in the mini-bar (0: No, 1: Yes)
SRQuietRoom	Boolean	Indication if the customer usually asks for a room away from the noise (0: No, 1: Yes)