

Data Description

This data article describes two datasets with hotel demand data. One of the hotels (H1) is a resort hotel and the other is a city hotel (H2). Both datasets share the same structure, with 31 variables describing the 40,060 observations of H1 and 79,330 observations of H2, we combined the two datasets into one file. Each observation represents a hotel booking. Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were canceled. Since this is hotel real data, all data elements pertaining hotel or customer identification were deleted.

Value of the data:

- Descriptive analytics can be employed to further understand patterns, trends, and anomalies in data.
- Used to perform research in different problems like bookings cancellation prediction, customer segmentation, customer satiation, seasonality, among others.
- Researchers can use the datasets to benchmark bookings' prediction cancellation models against results already known (e.g. [1]).
- Machine learning researchers can use the datasets for benchmarking the performance of different algorithms for solving the same type of problem (classification, segmentation, or other).
- Educators can use the datasets for machine learning classification or segmentation problems.
- Educators can use the datasets to obtain either statistics or data mining training.

Variable	Unique	Top counts
Agent	186	240: 13 095, NULL: 8 209, 250: 2 869, 241: 1 721
ArrivalDateMonth	12	Aug: 4 894, Jul: 4 573, Apr: 3 609, May: 3 559
AssignedRoomType	11	A: 17 046, D: 10 339, E: 5 638, C: 2 214
Company	236	NULL: 36 952, 223: 784, 281: 138, 154: 133
Country	125	PRT: 17 630, GBR: 6 814, ESP: 3 957, IRL: 2 166
CustomerType	4	Tra.: 30 209, Tra.-Party: 7 791, Con.: 1 776, Gro.:284
DepositType	3	No Dep.: 38 199, Non-Refund.: 1 719, Ref.: 142
DistributionChannel	4	TA/TO: 28 295, Dir.: 7 865, Cor.: 3 269, Und.: 1
IsCanceled	2	0: 28 938, 1: 11 122
IsRepeatedGuest	2	0: 38 282, 1: 1 778
MarketSegment	6	Onl.: 17 729, Off.: 7472, Dir.: 6 513, Gro.: 5 836
Meal	5	BB: 30 005, HB: 8 046, Und.: 1 169, FB: 754
ReservationStatus	3	C.Out: 28 938, Can.: 10 831, No-Show: 291
ReservedRoomType	10	A: 23 399, D: 7 433, E: 4 892, G: 1610

Variable	Unique	Top counts
Agent	224	9: 31 955, NULL: 8 131, 1: 7 137, 14: 3 640
ArrivalDateMonth	12	Aug: 8 983, May: 8 232, Jul: 8 088, Jun: 7 894
AssignedRoomType	9	A: 57 007, D: 14 983, E: 2 168, F: 2 018
Company	208	NULL: 75 641, 40: 924, 67: 267, 45: 250
Country	166	PRT: 30 960, FRA: 8 804, DEU: 6 084, GBR: 5315
CustomerType	4	Tra.:59 404, Tra.-P.: 17 333, Con.: 2 300, Gro.:293
DepositType	3	No Dep.: 66 442, Non-Refund.: 12 868, Ref.: 20
DistributionChannel	5	TA/TO: 68 945, Dir.: 6 780, Cor.: 3 408, GDS: 193
IsCanceled	2	0: 46 228, 1: 33 102
IsRepeatedGuest	2	0: 77 298, 1: 2 032
MarketSegment	8	Onl.: 38 748, Off.: 16 747, Gro.: 13 975, Dir.: 6 093
Meal	4	BB: 62 305, SC: 10 564, HB: 6 417, FB: 44
ReservationStatus	3	C.Out: 46 228, Can.: 32 186, No-Show: 916
ReservedRoomType	8	A: 62 595, D: 11768, F: 1 791, E: 1 553

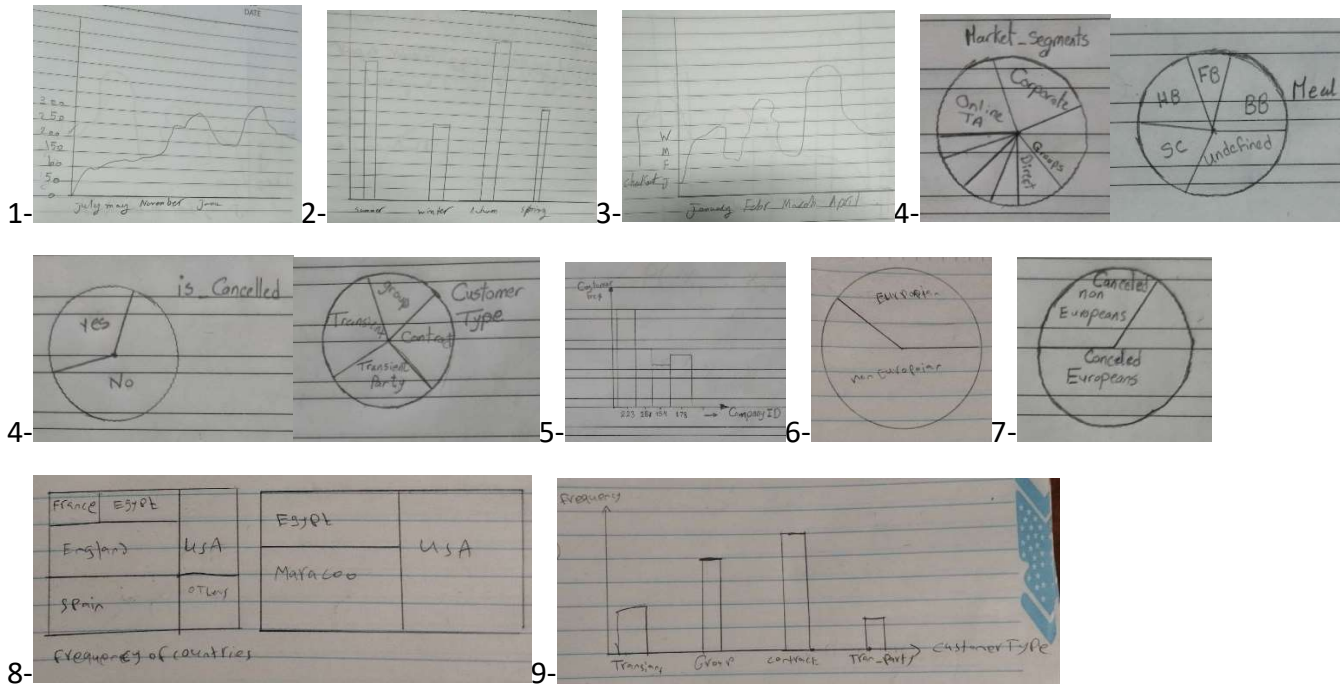
Variable	Type	Description			
ADR	Numeric	Average Daily Rate as defined by [5]	DistributionChannel	Categorical	Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
			IsCanceled	Categorical	Value indicating if the booking was canceled (1) or not (0)
Adults	Integer	Number of adults	IsRepeatedGuest	Categorical	Value indicating if the booking name was from a repeated guest (1) or not (0)
Agent	Categorical	ID of the travel agency that made the booking			
ArrivalDateDayOfMonth	Integer	Day of the month of the arrival date			
ArrivalDateMonth	Categorical	Month of arrival date with 12 categories: "January" to "December"			
ArrivalDateWeekNumber	Integer	Week number of the arrival date			
ArrivalDateYear	Integer	Year of arrival date	LeadTime	Integer	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
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	MarketSegment	Categorical	Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
Babies	Integer	Number of babies			
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	Meal	Categorical	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)
Children	Integer	Number of 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	PreviousBookingsNotCanceled	Integer	Number of previous bookings not cancelled by the customer prior to the current booking
Country	Categorical	Country of origin. Categories are represented in the ISO 3155–3:2013 format [6]			
			DaysInWaitingList	Integer	Number of days the booking was in the waiting list before it was confirmed to the customer
CustomerType	Categorical	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	DepositType	Categorical	Indication on if the customer made a deposit to guarantee the booking. This variable can 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.
PreviousCancellations	Integer	Number of previous bookings that were cancelled by the customer prior to the current booking	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
RequiredCardParkingSpaces	Integer	Number of car parking spaces required by the customer	ReservedRoomType	Categorical	Code of room type reserved. Code is presented instead of designation for anonymity reasons
ReservationStatus	Categorical	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	StaysInWeekendNights	Integer	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
			StaysInWeekNights	Integer	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
			TotalOfSpecialRequests	Integer	Number of special requests made by the customer (e.g. twin bed or high floor)

Questions:

1. Correlation between ADR & ArrivalDateMonth (**Line**)
2. Which season has the most reservations? (**Bar**)
3. (**Line**) chart between Month and Checked Out in Reservation Status
4. (**Pie**) Chart for Market Segments, Meal, percentage isCanceled and Customer Type

5. **(Bar)** Chart for Company against frequency of Customers
6. Percentage of Europeans VS non-Europeans **(Pie)**
7. Percentage of Canceled Europeans VS Canceled non-Europeans **(Pie)**
8. **(Tree)** Map for Countries according frequency of Customers and Countries that cancel the most
9. Which Customer Type cancels the most **(Bar)**

Pictures for each Question:



Tasks and Schedule:

We all cooperated to finish all the tasks

EDITED QUESTIONS:

Questions:

1. Correlation between ADR & ArrivalDateMonth **(Pie)**
2. Showing the months of the year and the arrival daily rate of customers and percentage of children arriving with them each month **(tree)**
3. Which season has the most reservations and what is the most season has huge cancelation? **(Bar)**
4. **(Line)** chart between Month and Checked Out in Reservation Status
5. comparison between each hotel and each type of meals **(interactive pie)**
6. **(Pie and donut)** Chart for Market Segments, Meal, percentage isCanceled and Customer Type
7. Which Customer Type and deposit type cancels the most and **(Stacked Bar)**
8. Percentage of Europeans VS non-Europeans **(Pie)**
9. Percentage of Canceled Europeans VS Canceled non-Europeans and what is most Cancellation Countries in each season **(sunburst)**
10. **(Tree)** Map for Countries according frequency of Customers and Countries that cancel the most
11. **(tree) What customer types have the most adults and the frequency of those customer types having children with them**
12. Checking how many adults come from which companies and how many come individually **(tree)**
13. word cloud for all countries (word cloud)

Edited

Tasks and Schedule:

Ragab Q 1,3,4

Diaa Q 6,7,13

Ziad Q 8,9,5

Yusuf Q 2,10,11,12