NOVA IMS
Information Management School

7

BUSINESS CASE 4: PREDICT HOTEL BOOKING CANCELLATIONS

Business Cases with Data Science

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Summary

1. Business situation
2. Key problems
3. Data
4. Assignment

7.1

Business situation

Case 4: Predict hotel booking cancellations

Introduction (1/2)

In the hotel industry, as in many other travel-related industries, demand is managed through advanced bookings. **Bookings (also known as reservations) are a forward contract between the hotel and the customer** that gives the customer the right to use the service in the future at a settled price, but often with an option to cancel (1).

The cancellation option puts the risk on hotels who have to honor the bookings that they have on-the-books, but, at the same time, have to support the opportunity costs of having vacant rooms when someone cancels, and there is no time to try to sell the room or sell it at a discounted price. In Europe, the cancellation rate by reservation value, from 2014 to 2018, rose from 33% to 40% (2).

NOVA IMS Information Management School

Introduction (2/2)

Concerned about the increasingly negative impact caused by cancellations, A, the Revenue Manager Director of hotel chain C, a chain with resort and city hotels in Portugal, hired a consultant to evaluate the possibility of developing predictive models to predict net demand for their hotels. The hotel provided the consultant datasets of one resort hotel (H1) and of one city hotel (H2), of bookings that were due to arrive between July 1, 2015, and August 31, 2017 (3).

The name of the individual and the company name is anonymized to protect confidentiality. The referenced data are real.

Net demand is defined as demand minus cancellations

NOVA IMS Information Management School

Background (1/4)

Cancellations occur for understandable reasons such as business meeting changes, vacation rescheduling, illness, or adverse weather conditions. However, cancellations also occur for not-so-understandable reasons, such as finding a better deal.

“Deal-seeking” customers tend to make multiple bookings for the same trip or make one booking but continue to search for better deals (e.g., looking for hotels with better social reputation, better price, or better locations) (4).

The number of “deal-seeking” customers has grown immensely with the appearance of Online Travel Agencies (OTAs) in 1996 (2, 5).

NOVA IMS Information Management School

Background (2/4)

Although OTAs provide high market exposure and the possibility of hotels for selling inventory at opaque and bundled rates, they also have negative aspects. OTAs charge commissions that range from 15% to 30%. OTAs force hotels to guarantee to them the best available price or force rate parity among the different distribution channels.

Due to the exposure that hotels gained with OTAs and online distribution, competition is fierce among hotels. This competition, together with OTAs push for hotels to practice a free cancellation policy (2), makes hotels employ controls such as overbooking to fight cancellations.

Background (3/4)

However, overbooking creates several problems:

- **Reallocation costs**: hotels must pay for the reallocation of customers to other hotels;
- **Social reputation damage**: customers sharing their unpleasant experience on social media, harm the hotel's social reputation;
- **Loss of immediate and future revenue**: hotels lose not only the revenue of the reallocated customer's current booking but also the possible future revenue of that customer, as probably, the customer will not want to book again at the hotel.

Background (4/4)

On the other side, restrictive cancellation policies, such as non-refundable rates, also create problems:

- **A decrease in revenue**: due to the discounts on prices;
- **A decrease in the number of bookings:** as most customers do not like these types of policies

7.2

Key problems

Case 4: Predict hotel booking cancellations

Cancellations in C hotel chain (1/3)

C hotel chain operation is not different from other independent and non-independent hotel chains. C hotel chain operation was severally impacted by cancellations, with cancellations, as presented in Table 1, representing almost 28% in H1 and almost 42% in H2.

For this reason, A, decided to limit the number of rooms sold with restrictive cancellation policies. To balance that decision, A implemented a more aggressive overbooking policy. However, the latter started to generate costs. To counterbalance those costs, A, soften the overbooking policy, which in turn was also revealed to be not good. The less aggressive overbooking policy resulted in the hotel having inventory not sold, even on high-demand dates.

Cancellations in C hotel chain (2/3)

To reduce the uncertainty about demand, A wanted to implement prediction models to allow the chain's hotels to forecast net demand based on reservations on-the-books.

With these models' estimations, A expected to implement better pricing and overbooking policies but also do identify bookings with a high likelihood of canceling. Identifying those bookings could allow the hotels to try to contact those bookings' customers and make offers to try to prevent cancellation (e.g., dinner, car parking, spa treatments, discounts, or other perks).

The goal of A was to reduce cancellations to a rate of 20%.

Cancellations in C hotel chain (3/3)

		Not Canceled	Canceled	Total
H1	Bookings	28 938 (72.2%)	11 122 (27.8%)	40 060 (100%)
	Room Revenue	11 601 850 € (66.5%)	5 842 177 € (33.5%)	17 444 028 € (100%)
H2	Bookings	46 228 (58.3%)	33 102 (41.7%)	79 330 (100%)
	Room Revenue	14 394 410 € (56.9%)	10 885 060 € (43.1%)	25 279 470 € (100%)

Table 1: Cancellations in H1 and H2

7.3

Data

Case 4: Predict hotel booking cancellations

NOVA IMS Information Management School

Data (1/7)

The provided dataset is composed of the following columns:

- **ADR**: Average Daily Rate
- **Adults**: Number of adults
- **Agent**: ID of the travel agency that made the booking
- **ArrivalDateDayOfMonth**: Day of the month of the arrival date
- **ArrivalDateMonth**: Month of arrival date with 12 categories: “January” to “December”
- **ArrivalDateWeekNumber**: Week number of the arrival date
- **ArrivalDateYear**: Year of the arrival date

Data (2/7)

- **AssignedRoomType**: 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
- **Babies**: Number of babies
- **BookingChanges**: 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
- **Children**: Number of children
- **Company**: ID of the company/entity that made the booking or is responsible for paying the booking. ID is presented instead of designation for anonymity reasons

Data (3/7)

- **Country**: Country of origin. Categories are represented in the ISO 3155-3:2013 format
- **CustomerType**: 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
- **DaysInWaitingList**: Number of days the booking was in the waiting list before it was confirmed to the customer

Data (4/7)

- **DepositType**: Indication 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 the stay.
- **DistributionChannel**: Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **IsCanceled**: Value indicating if the booking was canceled (1) or not (0)
- **IsRepeatedGuest**: Value indicating if the booking name was from a repeated guest (1) or not (0)

Data (5/7)

- **LeadTime**: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
- **MarketSegment**: Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **Meal**: 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)
- **PreviousBookingsNotCanceled**: Number of previous bookings not cancelled by the customer prior to the current booking

Data (6/7)

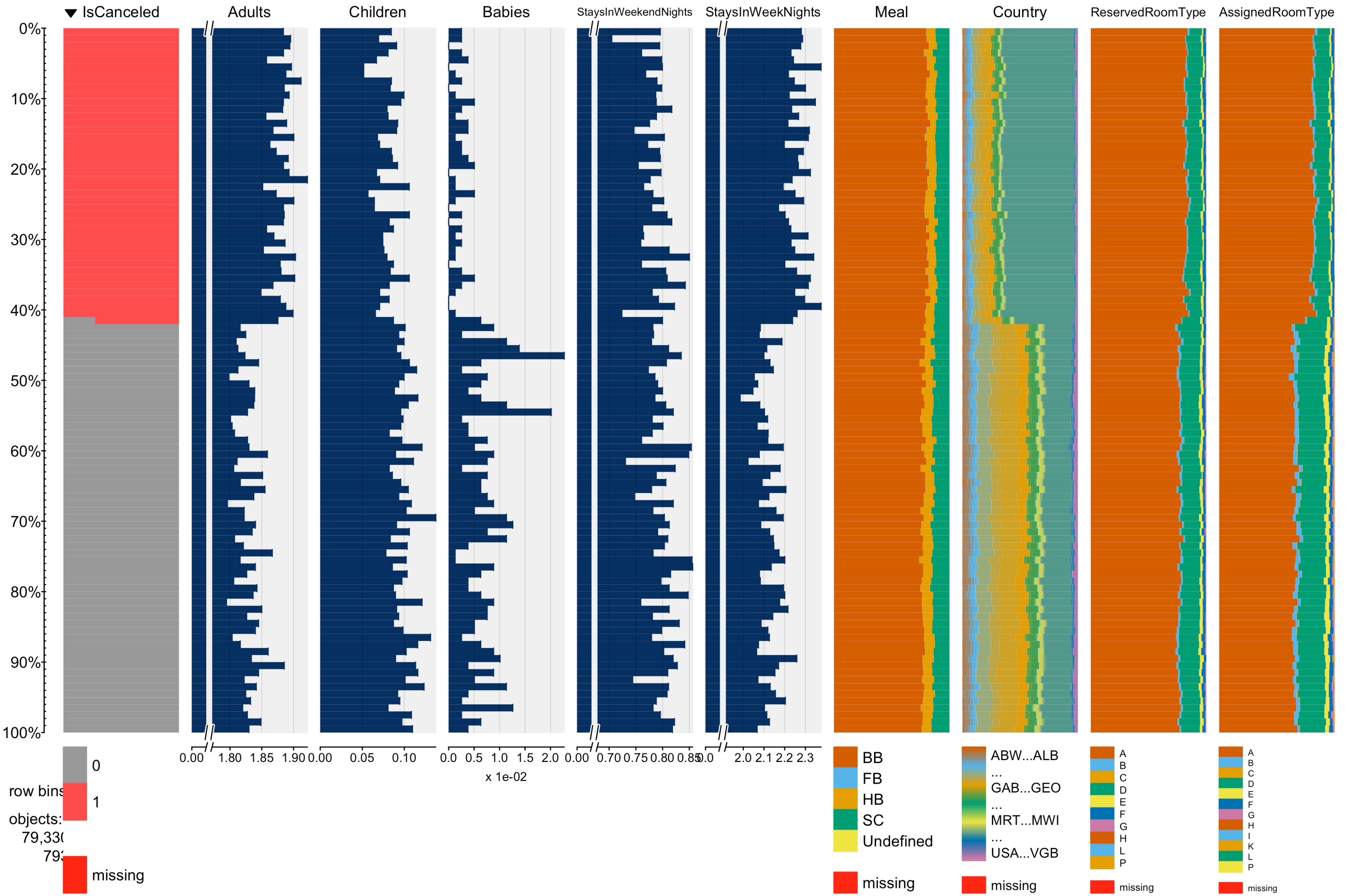
- **PreviousCancellations:** Number of previous bookings that were cancelled by the customer prior to the current booking
- **RequiredCarParkingSpaces:** Number of car parking spaces required by the customer
- **ReservationStatus:** Reservation last status, assuming one of three categories:
  - Canceled – 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

NOVA IMS Information Management School

Data (7/7)

- **ReservationStatusDate:** 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
- **ReservedRoomType:** Code of room type reserved. Code is presented instead of designation for anonymity reasons
- **StaysInWeekendNights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **StaysInWeekNights:** Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **TotalOfSpecialRequests:** Number of special requests made by the customer (e.g. twin bed or high floor)

H2 DATASET (PART OF COLUMNS)



NOVA IMS Information Management School

7.4

Assignment

Case 4: Predict hotel booking cancellations

Assignment (HOTEL H2)

1. Explore the data and build a model to predict cancellations:
    1. Define a machine learning success criteria
    2. Take into consideration the business objectives and requirements when selecting the algorithm
2. Elaborate on the business implications of employing the model and the insights obtained from model development
3. Make suggestions on how could the model be deployed and its impact on the hotel's business processes

Questions?

Business Cases with Data Science

NOVA IMS Information Management School

References

1. TALLURI, Kalyan T. and VAN RYZIN, Garrett. The theory and practice of revenue management. New York, NY : Springer, 2005. International series in operations research & management science, 68. ISBN 1-4020-7701-7.
2. HERTZFELD, Esther. Study: Cancellation rate at 40% as OTAs push free change policy. Hotel Management [online]. April 23 2019. [Accessed August 15 2019]. Available from: https://www.hotelmanagement.net/tech/study-cancelation-rate-at-40-as-otas-push-free-change-policy
3. ANTONIO, Nuno, ALMEIDA, Ana de and NUNES, Luis. Hotel booking demand datsets. Data in Brief. February 2019. Vol. 22, p. 41–49. DOI 10.1016/j.dib.2018.11.126.
4. ANTONIO, Nuno, DE ALMEIDA, Ana and NUNES, Luis. Big Data in hotel Revenue Management: Exploring cancellation drivers to gain insights Into booking cancellation behavior. Cornell Hospitality Quarterly. May 29 2019. P. 193896551985146. DOI 10.1177/1938965519851466.
5. BARTHEL, Jill and PERRET, Sophie. OTAs – A hotel’s friend or foe? [online]. London, UK : HVS, 2015. [Accessed August 15 2019]. Available from: https://www.hospitalitynet.org/file/152005663.pdf