

Agoda cancellation prediction

Home Test

Motivation

Understanding cancellation behavior is of great importance to Agoda's Finance Team, to our Product Owners who wish to make decisions in a short amount of time and to our partners, that don't have the luxury of handling hundreds of millions of booking records and understanding the cancellation patterns.

Machine learning can be used to create better forecasts for our net room nights at the booking level by predicting whether a guest will stay at the hotel or not. Knowing "when" the booking will cancel is also of importance, since we rarely ask the question of whether a booking will cancel on day 1 of their lifetime.

Terminology

Term	Description
Booking date	The date the customer purchases the booking
Checkin date	The date when the customer checks in the hotel, should not be earlier than the booking date
Checkout date	This is the end of the stay period. Must be at least 1 day later than checkin date.
Cancellation date	Optional – can occur any time from the booking date to the checkout date. Some customers cancel during the stay.

Objective

Based on the supplied data, predict if a set of bookings will cancel within the 11 days starting at 2018-12-07 and ending at 2018-12-13 (inclusive).

The training set consists of booking records along with a `cancellation_date` column that says when they cancelled.

If a booking record has a blank `cancellation_date` column it means the booking was not cancelled.

The booking records are for a sample of users, for those users we only selected bookings that:

- Have checkout date between 2018-07-02 and 2018-09-30
- Have booking date between 2017-07-02 and 2018-09-30

The test set consists booking that were:

- Booked between 2018-11-01 and 2018-11-30
- Have a checkin date strictly after 2018-12-15

Your goal is to predict which were cancelled between 2018-12-07 and 2018-12-13

We score the submission by Area Under the ROC Curve (AUC).

Data

The provided data is at booking level.

There are 58,659 records for learning.

There are 7,818 records for the test set.

Column	Description
h_booking_id	Hashed booking id
booking_datetime	Date and time the booking occurred
checkin_date	Date of checkin
checkout_date	Date of checkout
hotel_id	The id of the booked hotel
hotel_area_code	A random code representing the area the hotel is at (e.g: Gush Dan)
hotel_city_code	A random code representing the city where the hotel is at (e.g.: Tokyo)
hotel_country_name	ISO2 of the hotel's country
hotel_brand_code	A random code representing the hotel's brand (e.g: Ibis Hotels)
hotel_chain_code	A random code representing the hotel's chain (e.g.: Marriot)
hotel_live_date	When the hotel was first created in the system
hotel_star_rating	Hotel's star rating, between 0 and 5 by 0.5
accommodation_type_name	Hotel, Hostel, Apartment, etc.
charge_option	When does the customer need to pay
h_customer_id	Hashed customer_id
customer_nationality	Country name of the customer's passport
guest_is_not_the_customer	Boolean indicating whether the person who bought the bookings (customer) is not the person who stays (guest)
guest_nationality_country_name	Country name of the guest's passport
no_of_adults	
no_of_children	
no_of_extra_bed	
no_of_room	How many rooms were booked
origin_country_name	Name of the country inferred from the IP address
language	Language used at the website
original_selling_amount	Cost of booking for the customer
original_payment_method	
original_payment_type	
original_payment_currency	ISO3 of currency code the customer was charged in
is_user_logged_in	Was the user logged in when they made the booking?
cancellation_policy_code	See below
is_first_booking	Was this the first booking made by this customer?
request_nonesmoke	Did the user request a none smoking room?
request_latecheckin	Did the user request to check-in late?
request_highfloor	Did the user request a high floor?
request_largebed	Did the user request a large bed?
request_twinbeds	Did the user request a twin beds?
request_airport	Did the user request an airport transfer?
request_earlycheckin	Did the user request to check-in early?
cancellation_datetime	Optional: If Null, it means the booking was not cancelled, if not it is the date of cancellation.

Understanding the cancellation policy codes

Cancellation policy codes are written in a certain format. They are usually split into 2 conditions that are a breach of the cancellation policy:

1. Cancellation within **D** days before check in - this will be a breach of the cancellation policy.
2. "No show" - the customer does not cancel, but fails to show at the hotel.

Cancellation charges are applied either in percentage of the entire deal (**P**) or in prices per night (**N**)

General template

{days-before-checkin}**D**{charge}

Examples

"14D100P" - means 100% charge for cancellation within 14 days before check-in.

"14D1N" – means a 1-night charge for cancellation within 14 days before check-in.

Concatenated cases

The hotel can make more severe charges as the cancellation is nearer to the check-in. Example highlighted in red:

"130D1N_45D100P" - cancellation within 130 days before check-in will cost 1 night, 45 days before check-in will cost 100% of the cost.

No show charges template

"No show" charges are optional and might appear in the end of the cancellation code, example highlighted in red:

100D100P_100P - cancellation within 100 days before check-in will be charged 100% of the booking, no show will be charged 100% of the booking.