Hotel Booking Demand Prediction

Overview

Every year, more than 140 million bookings are made on the internet and many hotel bookings made through top-visited travel websites like Booking.com, Expedia.com, Hotels.com, etc. According to Google data, hotels are booked in advance of 12 weeks.

In this hackathon, we challenge data science enthusiasts to predict hotel booking cancellations using 33 distinguishing features.

It would be very helpful for the hotels to have such a model to decide if a guest will actually come or not. So that they can plan things like personel and food requirements in advance.

Data Dictionary

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Train File

CSV contains the booking details for whom the target variable **'is_canceled'** is known.

Variable	**Description**
Transaction_Id	Transaction Id
hotel	Hotel (H1 = Resort Hotel or H2 = City Hotel)
is_canceled	Value indicating if the booking was canceled (1) or not (0)
lead_time	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
arrival_date_year	Year of arrival date
arrival_date_month	Month of arrival date
arrival_date_week_number	Week number of year for arrival date
arrival_date_day_of_month	Day of arrival date
stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

stays_in_week_nights	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
adults	Number of adults
children	Number of children
babies	Number of babies
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)
country	Country of origin. Categories are represented in the ISO 3155–3:2013 format
market_segment	Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
distribution_channel	Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
is_repeated_guest	Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_cancellations	Number of previous bookings that were cancelled by the customer prior to the current booking
previous_bookings_not_canceled	Number of previous bookings not cancelled by the customer prior to the current booking
reserved_room_type	Code of room type reserved. Code is presented instead of designation for anonymity reasons
assigned_room_type	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
booking_changes	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
deposit_type	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.
agent	ID of the travel agency that made the booking
company	ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons
days_in_waiting_list	Number of days the booking was in the waiting list before it was confirmed to the customer
customer_type	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 asso-ciated 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 istransient, but is associated to at least other transient booking
adr	Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
required_car_parking_spaces	Number of car parking spaces required by the customer

total_of_special_requests	Number of special requests made by the customer (e.g. twin bed or high floor)
reservation_status	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
reservation_status_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

Test File

CSV containing the booking details for whom **'is_canceled'** is to be predicted.

Submission File Format

Variable	**Description**
Transaction_Id	Transaction Id
is_canceled	Predicted value indicating if the booking was canceled (1) or not (0)

Public and Private LeaderBoard

Test file is further divided into Public (25%) and Private (75%).

- Your initial responses will be checked and scored on the Public data.
- The final rankings would be based on your private score which will be published once the competition is over.

Evaluation Criteria

Your model performance will be evaluated on the basis of **ROC-AUC score**.

Rubrics

Component	Weightage
Data Cleaning and Data Visualization	25%
Model Building and Evaluation	60%
Pipeline and Deployment (Dashboard/Webapp)	15%