Hotel Booking Cancelation Classification

By: Sarah Alabdulwahab & Asma Althakafi

Introduction

Have you ever wondered if people tend to cancel their hotel bookings? Or to be exact, what are the features that affect that decision? This is our mission for this project. We aim to classify the customers that will cancel their hotel booking and those that will not.

Data Description

We obtained the dataset from Kaggle.com (from this link) and it contains 32 features and 119,390 hotel booking information. These are the features and their description:

- **hotel**: Resort Hotel or 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 weeknights (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 canceled by the customer prior to the current booking.
- **previous_bookings_not_canceled**: Number of previous bookings not canceled 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.

- o **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.
- o **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.
- o **deposit_type**: Indication on if the customer deposited 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.
- o **agent**: ID of the travel agency that made the booking.
- o **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.
- o **days_in_waiting_list**: Number of days the booking was on the waiting list before it was confirmed to the customer.
- o **customer_type**: Type of booking, assuming one of four categories:
  - Contract – when the booking has an allotment or other type of contract associated with it
  - Group – when the booking is associated with a group
  - Transient – when the booking is not part of a group or contract and is not associated with another transient booking
  - Transient-party – when the booking is transient but is associated with at least another transient booking
- o **ADR**: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
- o **required_car_parking_spaces**: Number of car parking spaces required by the customer.
- o **total_of_special_requests**: Number of special requests made by the customer (e.g. twin bed or high floor).
- o **reservation_status**: Reservation's 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 – the customer did not check-in and did inform the hotel of the reason why
- o **reservation_status_date**: Date at which the last status was set.

Our target variable is **"is_canceled"**.

Tools

- o Pandas and Numpy for data manipulation.
- o Sklearn for classification.
- o Matplotlib and Seaborn for plotting.
- o Tableau for interactive visualizations.

MVP Goal

The expected outcome is a classification of customers that will cancel their hotel booking and those that will not, a dataset, a report of the analysis, an interactive dashboard, and finally, a presentation highlighting all the main points from beginning to end.