

Project Title: **Predicting Hotel Bookings Cancellation**

Machine Learning Classification Model

By: Eman Alwusaybie

Project Description

Hotel Booking Demands It is an industry in itself. So, it is important to manage reservations very efficiently, based on modern technologies such as machine learning.

In this project, I will seek to predict the possibility of a booking for a hotel based on different factors and try to predict if they need special requests based on different features.

The data set contains booking information for a city hotel. It includes information such as when the booking was made, the, and the number of available parking spaces, Lead Time, Booking Changes, Previous Cancellations, among other things. From it, I can understand the customer's behavior and it might help me make better decisions.

Goals /Questions

- The aim of this project is to predict a hotel booking's likelihood to be canceled. The data includes hotel booking information of two hotels: a city hotel (H1) and resort hotel (H2) and total 31 features in the dataset.
- The model aims to understand bookings made from July 2015 onwards, and will attempt to predict based on booking characteristics, whether a customer has a higher chance of cancellation.
- This model will be useful in identifying potential customers who can cancel and help hotels apply appropriate cancellation policies depending on the marketing strategy

Dataset

I will use the Hotel Booking Demand dataset from the Kaggle with 119.390 rows and 31 columns.

You can download it from here: <https://www.kaggle.com/jessemostipak/hotel-booking-demand>

This data set contains booking information for a city 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. I have selected only the **17** most promising features contained in this dataset as shown below:

Variable	Description
adr	Average Daily Rate, calculated by dividing the sum of all lodging transactions by the total number of staying nights
adults	Number of adults
agent	ID of the travel agency that made the booking
arrival_date_week_number	Week number of the arrival date
booking_changes	Number of changes/amendments made to the booking from the moment the booking was entered on the Property Management System (PMS) until the moment of check-in or cancellation
country	Country of origin
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 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 another transient booking
hotel	Type of hotel
lead_time	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
market_segment	Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
previous_bookings_not_canceled	Number of previous bookings not canceled by the customer prior to the current booking
previous_cancellations	Number of previous bookings that were canceled by the customer prior to the current booking
required_car_parking_spaces	Number of car parking spaces required by the customer
reserved_room_type	Code of room type reserved. Code is presented instead of designation for anonymity reasons
stays_in_week_nights	Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
total_of_special_requests	Number of special requests made by the customer (e.g. twin bed or high floor)

Packages / Libraries

- Pandas
- Numpy
- Seaborn
- Matplotlib
- sklearn

Machine Learning Models

- Random Forest Classifier
- Logistic Regression