Hotel Booking Cancelation Prediction Proposal

There are a lot of people who need to stay in hotels for many reasons, including work or a pleasant vacation with family, and there are also many reasons that can influence the decision to stay at the hotel or cancel the booking. By analyzing the previous data of hotel bookings, we may predict if the customer will cancel or confirm a booking.

Question/need:

The purpose of the model I plan to build is to predict the possibility of canceling a hotel booking. This information can be useful for the hotel industry. Many hotels offer free cancelation option or payment upon arrival. It will be useful for hotel managers to know the probability of cancelation for each customer. This will allow them to make actions to increase the probability of booking confirmation, such as reminding or calling the customers with high probability of canceling first. In case of no response, they can make the booked room available if the customer is late a few hours, but they hold the room longer for those with low probability to cancel. The goal of these actions is to increase profits. The model I plan to build will help decision makers in the hotel industry to take these actions at the right time with the right customers.

Data Description:

- The dataset is hotel booking demand data downloaded from kaggle.com. The dataset contains 119390 samples and 32 features.

- The individual sample is user booking information for one customer.

- In this project, we need to use the most useful of them for analysis like (`'hotel', 'lead_time', 'arrival_date_year','arrival_date_month', 'arrival_date_week_number','arrival_date_day_of_month', 'stays_in_weekend_nights','stays_in_week_nights', 'adults', 'children','meal, 'market_segment', 'distribution_channel', 'is_repeated_guest', 'previous_cancellations','previous_bookings_not_canceled', 'reserved_room_type','assigned_room_type', 'booking_changes','deposit_type','company', 'days_in_waiting_list','customer_type', 'adr', 'required_car_parking_spaces', 'total_of_special_requests','reservation_status', 'reservation_status_date'`).

- The target I want to predict is (`is_canceled`). The value 1 indicates the booking was canceled and 0 means the booking was confirmed.

Tools:

- Python Language
- Pandas, NumPy for data processing
- Scikit-learn library for building the model
- Matplotlib for visualization.

MVP Goal:

The minimum viable product (MVP) would be a Jupyter Notebook document, where I state the goal of this project and show how I achieve it by using machine learning algorithm and including plots and my conclusions.