ORIE 5741 Project Proposal

We are writing a proposal to the manager of a hotel to pursue the following question: *Can we predict the likelihood of a hotel booking being canceled?* We will explore this question by analyzing a dataset of hotel reservation records from 2017 to 2018. Using data from this timeframe will ensure we can develop a model for a future that is more similar to the pre-pandemic era during which traveling and hotel booking were more frequent. Possible features for the input of the model are the number of guests, duration of stay, booking date, type of meal plan, whether the customer requires parking, the type of room reserved, lead time, arrival date, whether the booking was done online or offline, whether the guest is a repeat customer, number of bookings previously canceled, number of bookings *not* previously canceled, the average price per room, and number of special requests. The outcome variable is the booking status which would be either "Canceled" or "Not Canceled." Overall, there are 17 features and 36,275 records [1].

It is important we pursue this project because booking cancellations result in losses of profit for hotels. When a customer cancels a booking, there is a cost to the hotel if they are not able to find someone else to book the room in time. Moreover, the percentage of booking cancellations within total bookings has been increasing over time since the ease of booking in recent years means that people can book rooms without being sure of their plans [2]. Knowing the likelihood of a hotel booking being canceled can aid hotels with this issue since it might help them decide how to take action in advance, such as through pricing, advertising, or overbooking if they are expecting a room cancellation.

This dataset is well-suited for predicting hotel booking cancellations due to several key factors. First, the fields in the dataset are well-labeled and clearly defined, making it easy to understand and work with. Second, the dataset includes a binary classification of each booking as either canceled or not canceled, providing a clear target variable for predictive modeling. Third, the dataset contains a diverse range of features that are likely to be relevant to predicting cancellations, such as room type reserved, repeated guests, number of adults/children, booking channel, and previous cancellations. All these features would be commonly available in most hotels, as a result, it would be easier to recreate this dataset and train and test new models. Finally, the dataset is relatively large, with over 32,000 records, and includes data from two different hotels, allowing for robust analysis of cancellation patterns across different contexts. Altogether, these factors make the Hotel Reservations Classification Dataset a strong choice for predicting hotel booking cancellations.

[1] Ahsan Raza, 2023, "Hotel Reservations Dataset", Kaggle. [Online]. Available: https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset

[2] P. Delgado, "Cancellations shooting up: implications, costs and how to reduce them." Mirai.com. https://www.mirai.com/blog/cancellations-shooting-up-implications-costs-and-how-to-reduce-them/