Analytic Plan

DSE6211

Adeline Casali

**Objectives**

Hotel booking cancellation rates in the US have increased a whopping 33% from 2019 to 2022, resulting in a total of 20% of all bookings canceled in 2022 (SHR Group, 2023). This project aims to develop a model utilizing neural networks to predict the probability of cancellation for every new booking at ABC Hotels. ABC Hotels has provided data on 35,000 past customers' booking and cancellation behavior to fit the model. Within this data, cancellation rates have varied over time but overall have risen from 14.8% in 2017 to 36.7% in 2018 (Figure 1). This model will allow ABC Hotels to target bookings with a high risk of cancellation with additional advertisements, offers, and discounts in an effort to retain them as customers. Furthermore, this model will reduce uncertainty in demand management decisions, such as overbooking, and the need for restrictive cancellation policies, which often result in negative social impacts (Antonio et al., 2019).

A graph showing the growth of a number of years

Description automatically generated

*Figure 1: Cancellation rates over time, based on ABC Hotels data*

**Data**

The data provided by ABC Hotels contains 35,000 observations (bookings) over 17 variables (columns). Within the data, 11,878 bookings were cancellations, and 24,360 were not cancellations, for a total cancellation rate of 32.8%. The label, also known as the target or dependent variable, being investigated in this scenario is the booking status feature which states whether the reservation was canceled. The other features being used to fit the model are the number of children and adults, the number of weeknights and weekend nights, the type of meal plan, whether they required a car parking space, the room type, lead time, arrival date, market segment type, whether they were a repeat guest, the number of previous cancellations and bookings not canceled, the average price per room, and the number of special requests. The only column that will not be included in the model is the booking ID column, as the variable contains randomly assigned booking IDs that contain no practical or analytical value. The categorical variables such as meal plan, room type, market segment type, and booking status will be encoded into integers to prepare this data for use in the model. Additionally, dates will be separated into the month, day of the month, and day of the week to extract more information. Finally, the data will be explored to identify any outliers or missing values as well as any correlations between variables that might cause interference in the model.

**Outcomes and Impact**

In practice, information for new bookings will be fed into the model to calculate the probability of the customers canceling their booking. The model will produce the probability of cancellation for new bookings as a value between 0 and 1. In turn, this information can be passed on to marketing teams and used to provide customers with a high risk of cancellation with additional offers or advertisements in an effort to prevent their cancellation. As cancellations result in a considerable loss of income to hotel businesses, the investment in this model will result in a reduction in the loss of income suffered from cancellations for ABC Hotels.

**References**

António, N., Almeida, A. de, & Nunes, L. (2019). Predictive models for hotel booking cancellation: A semi-automated analysis of the literature. *Tourism & Management Studies*, *15*(1), 7–21. https://doi.org/10.18089/tms.2019.15011

*We need to talk about cancellations*. SHR Group. (2023, October 13). https://shrgroup.com/2023/06/21/we-need-to-talk-about-cancellations/