Hotel Reservations Analysis

MKT 6337.003 - Predictive Analytics for Data Science - S25

Group 6

Choksi, Ria A
Jana, Sunayana
Parida, Priyadarshan
Qazi, Khurem
Rajappan, Soniya

THE UNIVERSITY OF TEXAS AT DALLAS

Research Question

a. *What novel question(s) do you aim to answer in your project?*
   - What factors contribute most significantly to hotel reservation cancellations?
   - Can customer and booking characteristics predict the likelihood of a cancellation?
   - What insights can we draw to make better pricing and promotion decisions?

b. *Who will be your potential audience?*
   - Hotel operations managers, marketing managers, finance officers, and executives in the hospitality industry.

c. *Why would they be interested in your question(s)?*
   - Marketing managers can use insights to implement targeted strategies to reduce cancellations, such as offering incentives or improving customer satisfaction.
   - Operations managers can optimize hotel occupancy with the available prediction.
   - Finance Team: Improved forecasting of occupancy rates and optimized pricing strategies.

Data

a. *What is the data source?* ScienceDirect - Hotel Bookings | Data

b. *What is the data period?* Data period: July 2015 - August 2017

c. *What is the level of observation?* Level of observation: Individual hotel booking

d. *Do you have repeated observations for a given party?* Yes, there are repeated observations for customers who are repeated guests or have multiple bookings during the data period.

Model

a. *What is the outcome of interest (Y variable)?* Outcome: is_canceled (binary variable indicating whether a booking was canceled).

b. *What covariates or predictors (X variables) do you plan to include in your model?*
   - *Customer demographics:* adults, children, babies | *Booking details:* lead_time, stays_in_weekend_nights, stays_in_week_nights, reserved_room_type, assigned_room_type, booking_changes, deposit_type, days_in_waiting_list, total_of_special_requests | *Market and distribution:* market_segment, distribution_channel, country | *Historical behavior:* is_repeated_guest, previous_cancellations, previous_bookings_not_canceled | *Pricing:* adr (Average Daily Rate)

c. *What statistical model(s) do you plan on using?* Logistic Regression, Decision Tree and Random Forest, Gradient Boosting (e.g., XGBoost)

Results/Conclusions

a. *What are the final results and conclusions?*
   - Identify key factors driving cancellations and provide a predictive model with high accuracy.
   - Insights will help hotels implement strategies to reduce cancellations, such as adjusting lead times, flexible deposit policies, or targeting specific customer segments with promotions.

b. *What conclusions can you expect, or reach based on your analyses?*
   - Customers with longer lead times, higher ADR, or no deposit are more likely to cancel.
   - Repeated guests and those with special requests are less likely to cancel.
   - Predictive models will enable hotels to manage cancellations and improve operational efficiency proactively.