Stat 5650 Project Proposal

Group Members

Michael Huber
Greg Hoffmann
Kristen Sohm
Varsha Reddy Mandadi

Description of the Data

Link to the data: https://www.kaggle.com/jessemostipak/hotel-booking-demand

Hotel Booking Demand

"This data set contains booking information for a city hotel and a resort 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. All personally identifying information has been removed from the data." — Description taken from Kaggle webpage.

The main response variable for this data set ("Is_canceled") is coded 0 or 1, where 1 indicates the booking was cancelled. The "Hotel" variable specifies between the resort hotel and the city hotel.

More details:

- Number of columns: 32
- Number of rows: 119391
- Dates covered: 2015-07-01 to 2017-08-31
- Column names (descriptions for the columns can be found on Kaggle):
    - Hotel
    - Is_canceled
    - 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
    - Babies
    - Meal
    - Country
    - Market_segment
    - Distribution_channel

    - Is_repeated_guest
    - Previous_cancellations
    - Previous_bookings_not_conceled
    - Reserved_room_type
    - Assigned_room_type
    - Booking_changes
    - Deposit_type
    - Agent
    - Company
    - Days_in_waiting_list
    - Customer_type
    - Adr (average daily rate)
    - Required_car_parking_spaces
    - Total_ofspecial_requests
    - Reservation_status
    - reservation_status_date

Questions to be Addressed

- Can we create a model that will predict whether or not a customer will cancel their hotel reservation?
- Can we predict whether a customer will be a no-show instead of canceling?
- Can we accurately predict whether a customer will book at the city hotel or the resort hotel based on their profile?
- Similar to the Nest data from class, can we treat the two types of hotel (city and resort) the same, or are they different from one another?
- Can we predict the peak time of the year for maximum bookings?
- Can we predict what type of customers or market segments contribute to the most bookings?

Methods to be Applied

- We will apply some of the prediction methods covered in class to see how they compare against one another for this data set.
- Methods to attempt:
    - Linear Discriminant Analysis (LDA)
    - Quadratic Discriminant Analysis (QDA)
    - Regression
    - Logistic Regression
    - K Nearest Neighbors (KNN)
    - Classification Trees