Project Proposal

LAM-duh: Xuliang Deng, Leah Okamura, Megan Richards

Oct 28th, 2021

Introduction

We are interested in studying the characteristics of hotel room reservations in the United States - particularly looking at predictors of the price of a hotel reservation. There are several significant points of relevance for understanding these relationships: understanding predictors of room cost could be used to help identify where new hotels could be successfully created, allow travelers to plan financially for future travel.

Generally, we are looking to use linear models to understand the contributing factors to hotel room price, as well as identify the strongest predictors. We hypothesize that a model with predictors of hotel type, reserved room type, assigned room type, company, meal, number of adults/children/babies, the average daily rate or daily cost, and the reservation status, will be a statistically significant predictor of hotel room price, and that the predictors will be significant except for company and number of adults/children/babies.

Data Description

The source of the dataset is Tiny Tuesday, https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-11/readme.md. This data set comes from an open hotel booking demand dataset from Antonio, Almeida and Nunes, 2019. It is sourced from this study https://www.sciencedirect.com/science/article/pii/S2352340918315191#f0010. Due to the dataset being over 100,000 observations, we have limited the observations to be only hotels from the US. The general characteristics being measured in the data are the different aspects of booking and staying at a hotel. For example, out of the 32 variables, some of the ones we find great interest in are hotel type, reserved room type, assigned room type, company, meal, number of adults/children/babies, the average daily rate or daily cost, and the reservation status.

Analysis Approach

In this section, you will provide a brief overview of your analysis approach. This includes:

Description of the response variable. Visualization and summary statistics for the response variable. List of variables that will be considered as predictors Regression model technique (multiple linear regression and logistic regression)

The response variable is adr, average daily rate which is define as by dividing the sum of all lodging transactions by the total number of staying nights. We are going to use hotel (Resort Hotel or City Hotel), reserved_room_type, assigned_room_type, company, meal(type of meal), reservation_status. For example, if the reservation_status is canceled or no-show, the price can be expected to be cheap because that is an indicator of hotel being not popular and the room is also cheap, which does not stop customoers from cancelling. Meal plan and hotel type (Resort Hotel or City Hotel) can indicate how much hotel could cost. We can also use reserved_room_type or assigned_room_type

A research question we are interested in is how do factors such as type of hotel and type of guest affect the average daily rate for a hotel. We would also be interested in seeing how stays in the weekend or the

weekday may affect the average daily rate for a hotel, and if they differ between the two hotel types, City and Resort hotels.

weekend or weekday, number of hotel guests,

Date Dictionary

hotel (character) - Hotel (H1 = Resort Hotel or H2 = City Hotel)

is_canceled (double) - Value indicating if the booking was canceled (1) or not (0)

lead_time (double) - Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

arrival_date_year (double) - Year of arrival date

arrival_date_month (character) - Month of arrival date

arrival_date_week_number (double) - Week number of year for arrival date

arrival_date_day_of_month (double) - Day of arrival date

stays_in_weekend_nights (double) - Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

stays_in_week_nights (double) - Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

adults (double) - Number of adults

children (double) - Number of children

babies (double) - Number of babies

meal (character) - Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)

market_segment (character) - Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”

distribution_channel (character) - Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”

is_repeated_guest (double) - Value indicating if the booking name was from a repeated guest (1) or not (0)

previous_cancellations (double) - Number of previous bookings that were cancelled by the customer prior to the current booking

previous_bookings_not_canceled (double) - Number of previous bookings not cancelled by the customer prior to the current booking

reserved_room_type (character) - Code of room type reserved. Code is presented instead of designation for anonymity reasons

assigned_room_type (character) - Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons booking_changes (double) - Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

deposit_type (character) - Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.

agent (character) - ID of the travel agency that made the booking

company (character) - ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons

days_in_waiting_list (double) - Number of days the booking was in the waiting list before it was confirmed to the customer

customer_type (character) - Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking

adr (double) - Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

required_car_parking_spaces (double) - Number of car parking spaces required by the customer

total_of_special_requests (double) - Number of special requests made by the customer (e.g. twin bed or high floor)

reservation_status (character) - Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – customer did not check-in and did inform the hotel of the reason why

reservation_status_date (double) - Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel