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,