**Exploratory Data Analysis on Cancelations of Hotel Reservations**

Blaine Bristow

Department of Data Science, Bellevue University

DSC 530: Data Exploration and Analysis

Professor Matthew Metzger

March 4, 2023

The main question of interest in this exploratory analysis was: do the number of adults, number of children, number of weeknights, number of weekend nights, and price have a correlation with whether the reservation is cancelled?

From the multiple regression performed, it appears that number of adults, number of weeknights, number of weekend nights, and price all have a positive correlation with the reservation being cancelled, while the number of children has a negative correlation.

An assumption was made that neither number of days between booking and stay not date of stay would have an effect on cancellation. These 2 variables were included in the dataset, so further analysis could be performed to determine whether this is true. Another variable included that may provide valuable insight is whether the customer is a repeat customer, as well as how many bookings they fulfilled and how many they cancelled.

More information regarding location of the hotel may have been helpful in further analysis. If hotels from many different climates were sampled, perhaps seasonal trends could be observed, such as high cancellation rates for hotels in New England from November through February.

I had assumed when performing the correlation test that an increase in nights reserved would result in an across the board decrease in price of the room. In reality, cheap rooms tended to only be used for a couple nights, so an increase in nights saw both an increase and decrease in price as a median price was approached.

I did struggle with figuring out how to manually set limits on the axes of plots generated using the “thinkplot” package. I wound up figuring out what I was doing wrong that the automatic axes were way off. Otherwise, I understood everything discussed this course, although I feel like my understanding of the concepts covered in the later chapters (null hypothesis tests and regression analysis) could be better.