IST 687 M005 – Final Project Report

# Hotel Cancellations Analysis Report

Elizabeth Jones, Tia Jones, Chukwukadibia Soronnadi, Lagarthucin K Legrand

# Introduction

The dataset we analyzed was pulled into R from a csv file and contained several columns of data to analyze. The data received consisted of 20 variables and 40060 observations. The goal while viewing this data was to recognize patterns that could predict drivers for why some hotel reservations become canceled and to consider ways our client may reduce cancellations based on those predictions.

# Business Questions

While viewing the data we aimed to understand the following questions:

- Where do the most reservations come from?
- What type of person cancels?
- What type of person doesn't cancel?
- Do attributes exist that may entice customers not to cancel?

# Data Acquisition, Cleansing, Transformation and Munging

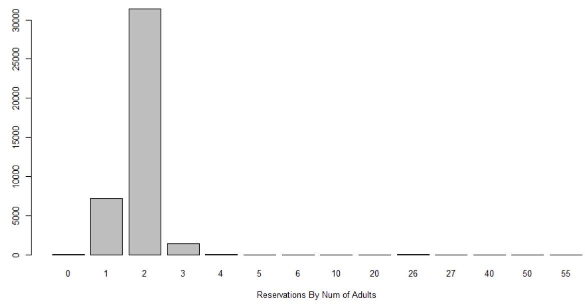
Within the dataset we pulled, there were no NA's within any of the columns with a numeric data type. However, the dataset did contain NULLs. We found that there were 432 reservations that had "NULL" for the country name. We decided to keep these rows of information for all our data analysis except for when we needed to specifically understand what country had the most reservations.

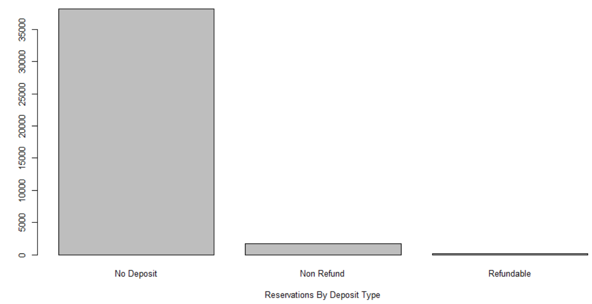
We created a few additional datasets in order to focus our analysis on specific results. For example, we needed to look at the difference between canceled and non-canceled reservations, so we split that information into two different datasets. We also created new datasets for our models that looked at specific columns as factors instead of integers or characters. Additionally, to view reservation percentages by country, we had to create a data set with those statistics. We then used that dataset to create a world map visualization that highlighted countries based on percentage of overall reservations.

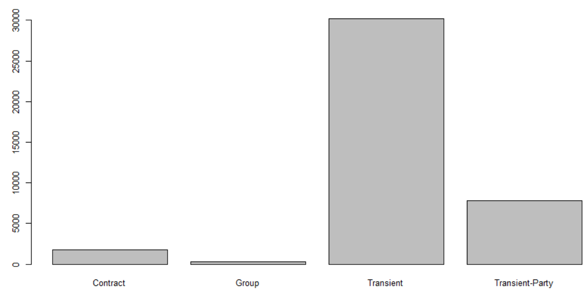
Finally, we added a column titled "Room Change" to compare whether reserved room type and assigned room type changed. True represents that Assigned Room Type and Reserved Room Type match.

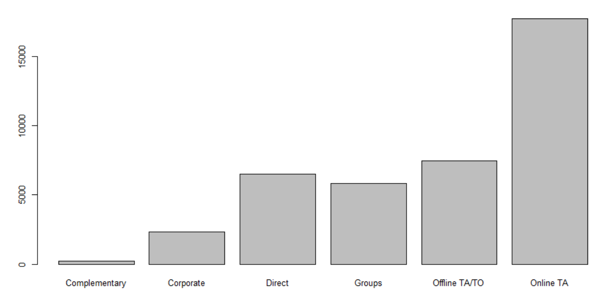
# Descriptive Statistics and Visualizations

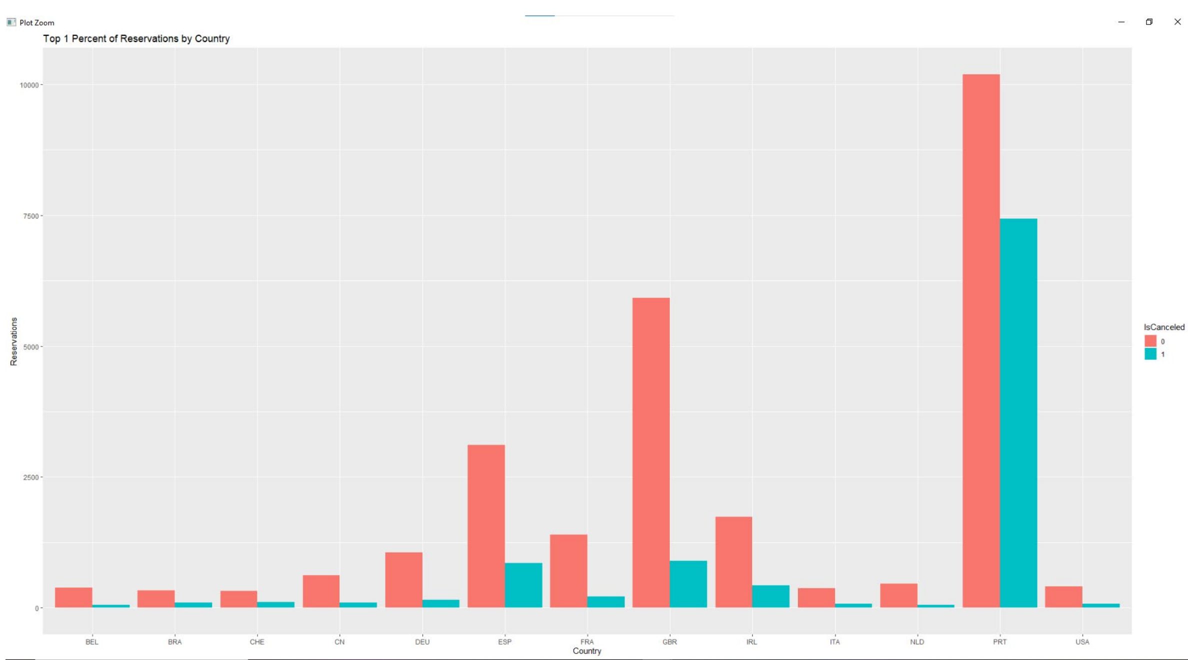
When we explored the data, we wanted to first focus on where we see the most reservations. We looked at Adults, Deposit Type, Customer Type, Market Segment and Country. We found that many reservations were coming from these markers 2 adults, No deposit, Transient, Online Travel Agent and Portugal.

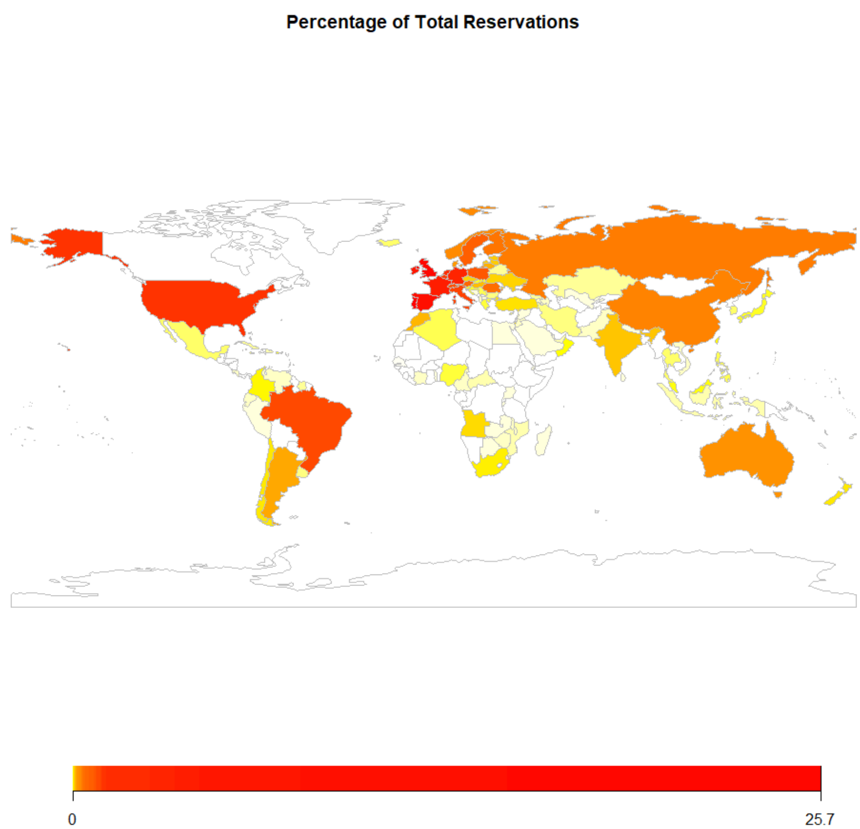


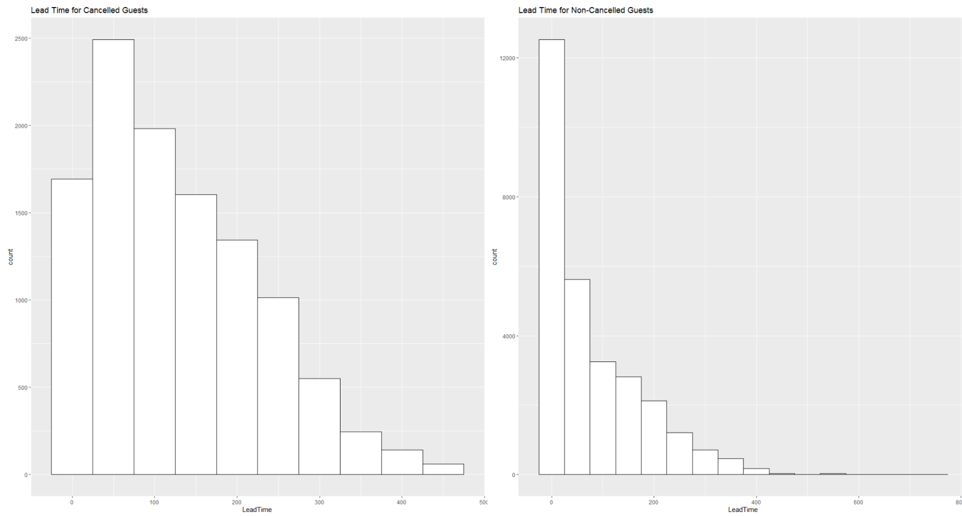



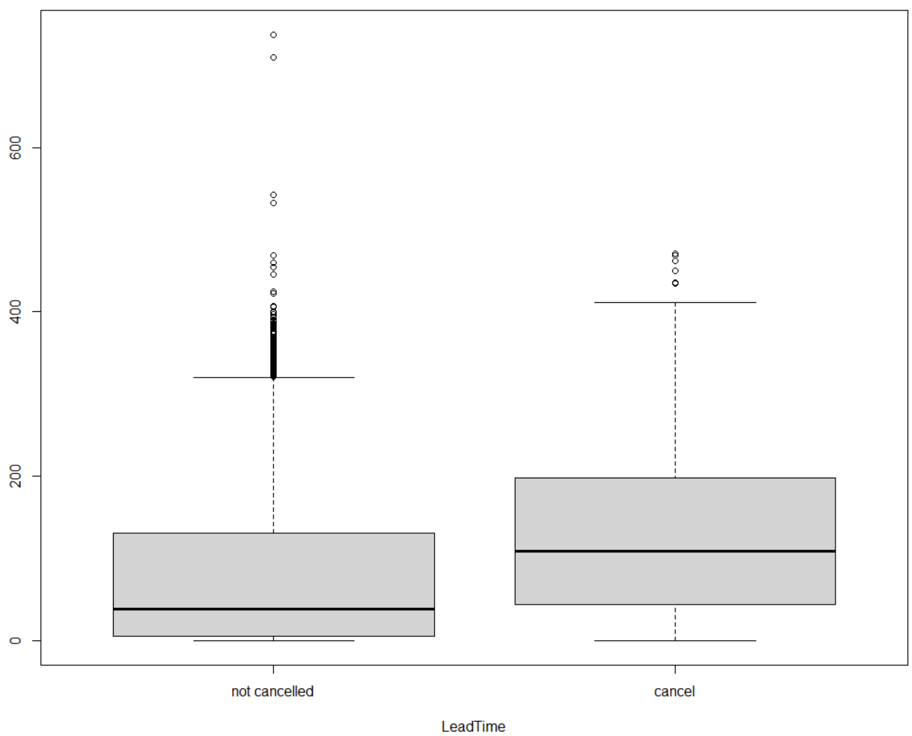



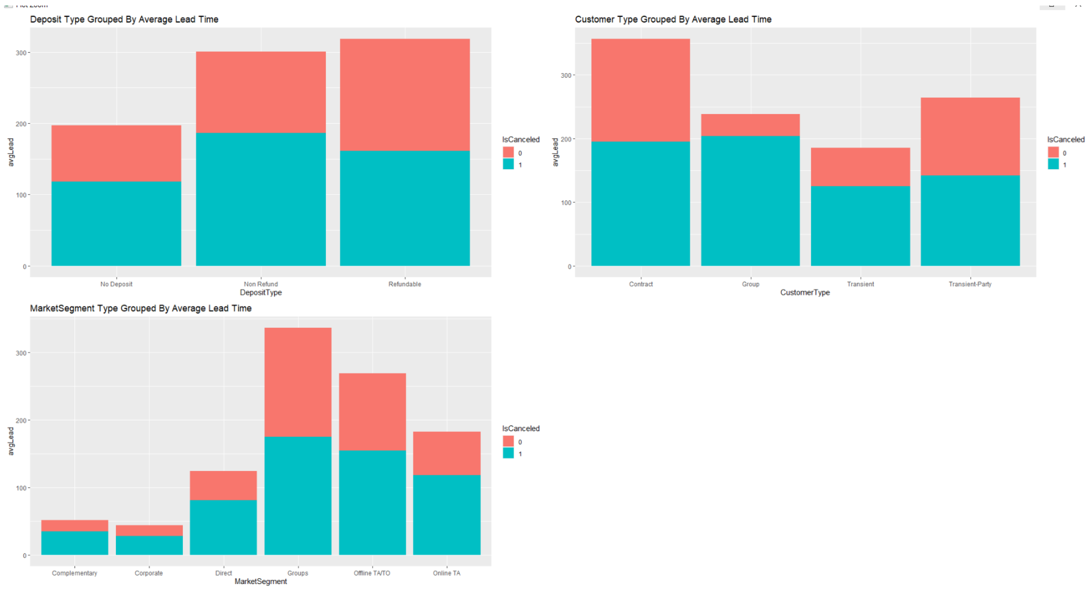



Next, we analyzed the Lead Time column as it is logical to assume that the shorter the lead time the least likely a person will cancel. When we compared the means of lead times between cancellations and non-cancellations there was indeed lower lead times with non-cancellations. We also checked distribution of Lead Times by frequency and percentage to better understand how the days spread out among reservations. When doing this we discovered that there were higher instances of cancellations when the lead time was a larger number than when it was smaller number comparatively.



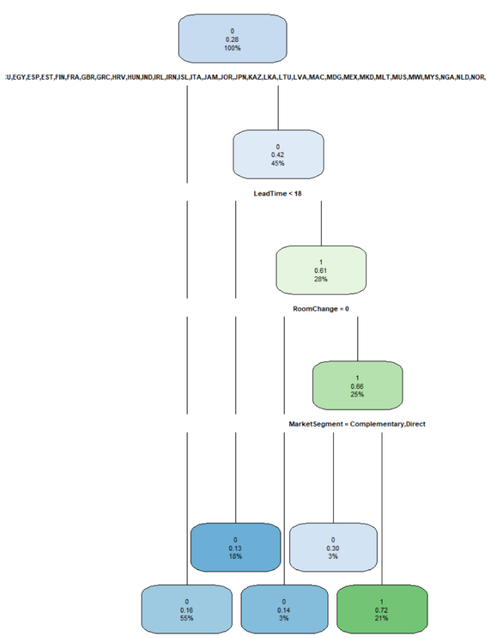




Last, we compared Deposit Type, Market Segment, and Customer Type with the average Lead Time to derive some meaning. We found that overall, fewer cancellations occurred when the lead times were lower. Based on this information it was probable to assume that Lead Time was a possible marker for cancellations.



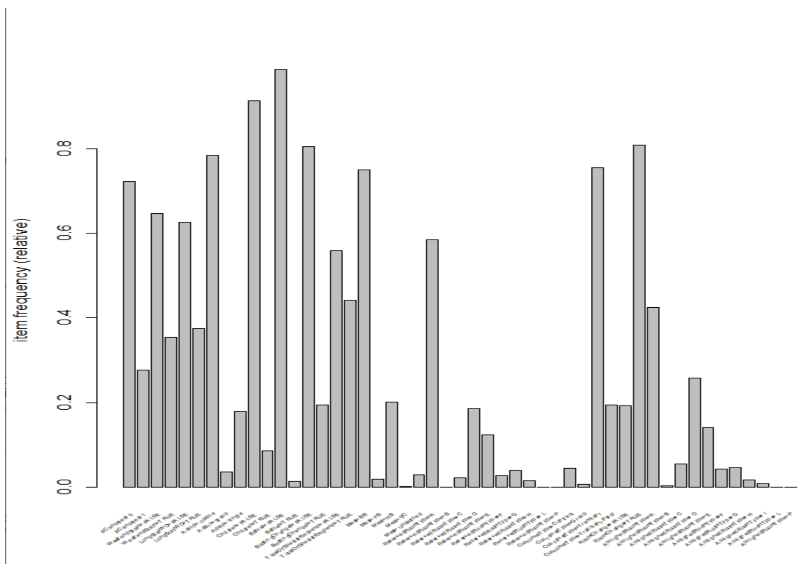

## Use of Modeling Techniques and Visualizations

Based on the initial observations we conducted a tree model to examine whether Lead Time was a significant marker of cancellations over others. For the tree model we looked at Country Type, Lead Time, Room Changes and Market Segment to predict whether someone might cancel. The model we created was 81% accurate and showed Lead Time as the strongest predictor followed by Market Segment and Room Changes.

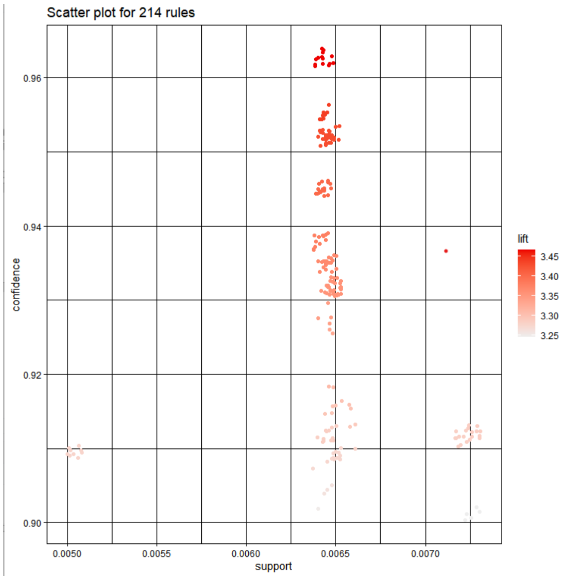


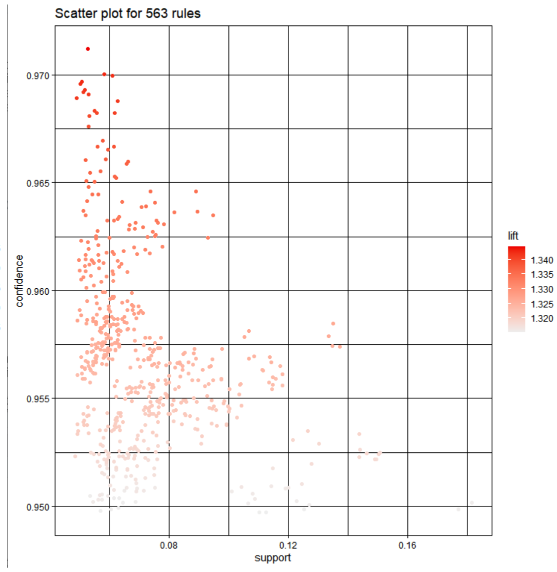
Next, we viewed predictions using the SVM model. With the SVM model we reviewed the same columns from the above previous model (Country Type, Lead Time, Room Changes, and Market Segment) to see if we would have similar accuracies. We chose a cost parameter of 5 and a cross validation of 3 to allow for a more specialized model. Once complete it showed a prediction accuracy of 83% which was slightly better than the prediction made by the tree model.

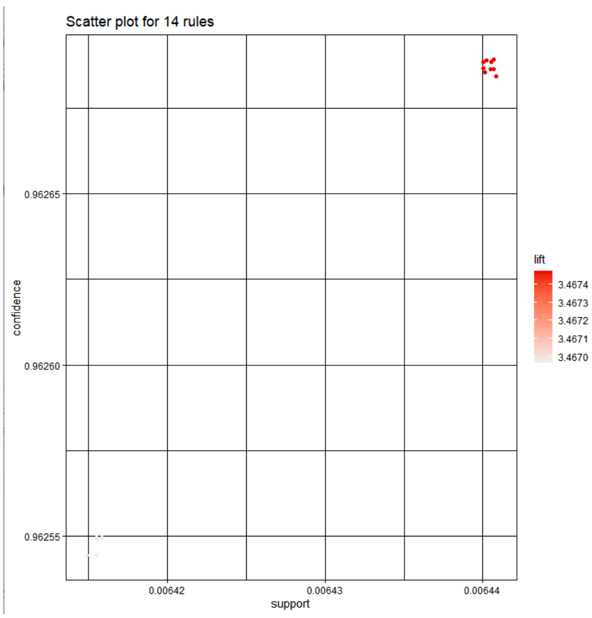
The third model used was the Association model. This model gave us an idea of any additional variables that impacted hotel cancellation rates in addition to Lead Time. First, we created and plotted frequencies using the ItemFrequency( ) to get a better idea of the item frequency distribution of each factor in our dataset.

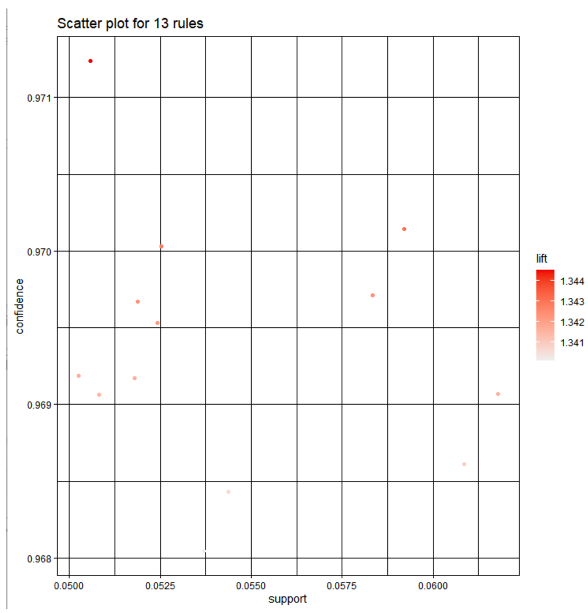



Next, we used the apriori( ) function to get a list of rules for customers that cancelled (IsCanceled=1) and customers that did not cancel (IsCanceled=0). Customers that canceled output over 200 rules, and customers that did not cancel output over 500 rules. We narrowed these lists down by selecting lists with high Confidence of at least 90%, high Support, so that we were picking attributes that showed up frequently in the dataset, and Lift with values greater than 1. We were left with 14 Rules for IsCanceled=1 and 13 Rules for IsCanceled=0.




After analyzing each list, we found that customers who cancelled were transient couples with the FB (Full board) meal option who reserved or were assigned to Room A. Those who did not cancel had the BB (Bed & Breakfast) meal option, no room changes, and had special requests.

# Actionable Insights

Based on our analysis, it appears that customers who cancel typically have larger lead times, no special requests and receive full breakfast with their booking. Whereas customers who did not cancel tend to consist of 2 adults who make special requests and book a room with a bed and breakfast package.

It would appear then that those who don't cancel are couples who may be planning getaway trips thus we would suggest additional marketing for special room packages to attract more of these types of customers.

To prevent customers from cancelling we would suggest putting a greater focus on advertising hotel services that would entice customers to keep their reservations. This could include things like giving customers the opportunity to purchase add-on services at discounted or special rates after making their initial booking.

# Appendix

Link to Descriptive, and Rpart Code: https://tinyurl.com/2p8ca267

Link to SVM Model Code: https://tinyurl.com/yc5593tt

Link to Association Model Code: https://tinyurl.com/2p8jha8b