# Midterm Project - Peer Review

by Mark Hageman for Eunsun Yook

November 5, 2020

Hi Eunsun! I really enjoyed reading your project so far! This looks like an interesting data set, and I think you are right about how helpful a predictive model like yours would be for hotel management. I had no idea that the percentage of cancelations was as high as it seems to be in this data set!

I did have one problem getting the .Rmd file to knit at first... An error was being generated at line 106 where you wanted to change the value of the hotels\_new$meal field to "SC" where it was "Undefined". I changed the line to read like this:

hotels\_new$meal[hotels\_new$meal == "Undefined"] <- "SC" #levels(hotels\_new$meal)[5] <- "SC"

This seemed to fix the issue. I just commented out your original code, as you can see.

Here is some feedback:

* I really like the navigation system, with the pills at the top! I had not considered doing it that way... I would like to try to do that with my project as well.
* In the Exploratory Data Analysis section, on the bar chart of market\_segment, a couple of the column headers are being truncated... Is it possible to widen the graph a bit to fit headers?
* With all the different categorical variables, there are lots of ways to break down the data. I think you have done a very thorough analysis of the different categories. I will be interested to see which classification model type does the best job for you.
* Another thing you did that I would like to copy in my report is the way you used the car::Boxplot() method to identify the outlier in the adr field, and its value. That was something I've tried to do before, but I did not know about the car package's Boxplot method. I think being able to label outliers with their values like you did is very informative.
* I thought it was interesting how higher lead\_time values seem to correlate positively with canceled bookings... In other words, it appears that for the most part, the bookings that got canceled tended to have been booked more in advance that the bookings that were *not* canceled.
* My partner Johnny and I are doing something similar with the description of the columns in the data set. We used the htmltools package, which allows you to include an external HTML file. We put the descriptions of the fields in the data set into an external file named "codebook.html" which helped us to organize our code quite a bit.

Good luck with the rest of the project! I think you're off to a great start!