datamemo

2022-10-03

Overview of dataset - The dataset includes a lot of important information about determining when hotel booking is really popular, which contains a lot of nonnumerical and numerical variables, such at the time of booking and how for how many people. - I would be obtaining the dataset from Kaggle and I could just download it from there. https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand (https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand) - There are around 119,390 observations and about 32 predictors. - I would be working with a mix of categorical, numerical, binary, and non-categorical variables. - There are missing data, like a lot for certain variables like reservation status and if a company made the booking. I plan to maybe ignore them since it isn't really relevant to what I'm looking for, but if it a big problem, I may need to exclude them or not use the variable.

Overview of research question(s): - I am interested in predicting what time of the year is when people begin to book hotels. My question is what is the busiest time for hotel bookings? - My response variables are the total number (count) of day (arrival_date_day_of_month), week (arrival_date_week_number), month (arrival_date_month), and year (arrival_date_year) of the time of the hotel booking. - My question would be best answered with a classification approach. - I think the predictors are the date the person booked their hotel. - It would be of a descriptive and inferential approach. I already have the data at hand so I am not predicting anything. I am trying to analyze when certain people book hotels at certain time (inferential) and also seeing the visual relationship between the response and predictor values,

Proposed project timeline: - I plan to have my data set loaded by next week. I will begin the writing specific questions that could be responded by this dataset by week 3. Week 4-5 should be cleaning the data. Week 6-8 would be focusing on exploratory data analysis. Week 9 would be making charts and graphs. Week 10 would be revisions and making sure everything works well.

Any questions and concerns: - I needed help cleaning up the data and dealing with variables that have a lot of missing values. No other questions.