Mark Hoyt

1/28/24

DSE 6211

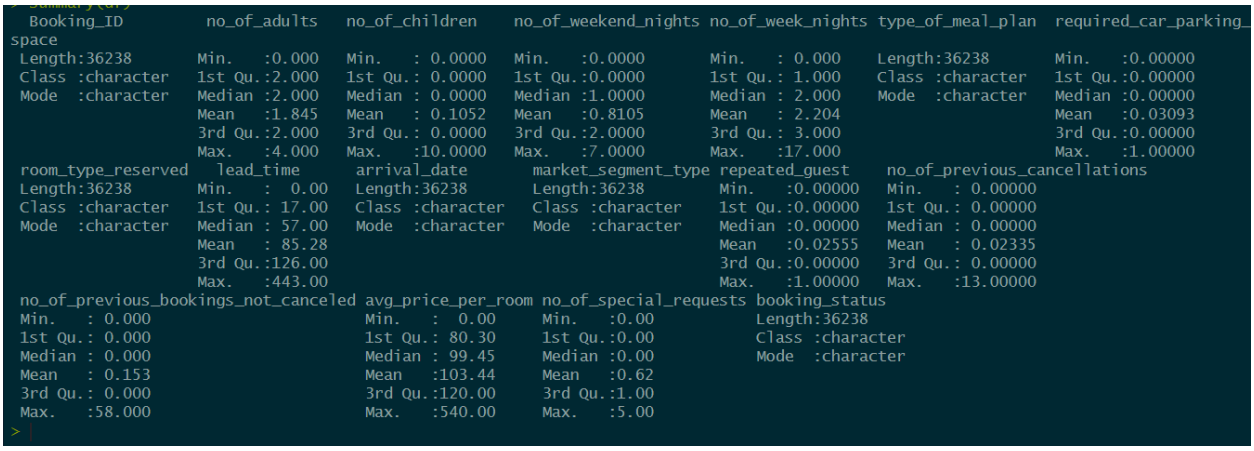
- Using the business need and project data available in Canvas (on the "Course Project" page of the "Start Here" module), propose a supervised classification problem to address the business need.
    - What is the label (i.e., the target or dependent variable) for the supervised classification problem? What will the model predict? How will they use the model to address their business need?

ABC Hotels is trying to identify bookings that have a high risk of cancellation. We have a variable of 0 or 1. 0 being the customer did not cancel and 1 being the customer did cancel, which gives us our dependent variable of booking_status. This model will help ABC hotels understand how other independent variables influence the likelihood of a customer canceling their reservation.

    - What data processing is needed and how will it be performed? Note: all variables should be included in the analysis unless a reason is given for exclusion. One column should be excluded. Any dates need to be processed.

I will be leveraging a Feedforward Neural Network to process the data and predict whether or not a customer is likely to cancel. I am using this model because it works well with regression and classification-supervised learning. Another form of processing I will be conducting is the standardization of columns and removal of outliers which will be identified in the data processing stage of this project. Dates in particular will be processed as we do not want a build-up on a given date or any replication errors. The column I will be excluding is booking_id as this is not

necessary for our business objective of predicting what type of customer is most likely to cancel.



- What features will be initially included? Some variables are as is. and some need to be extracted. month and date can be featured.

Featured columns: "no_of_adults", "no_of_children", "room_type_reserved", "lead_time", "arrival_date", "market_segment_type", "repeated_guest", "no_of_previous_cancellations", "no_of_previous_bookings_not_canceled", "avg_price_per_room", "no_of_special_requests", "booking_status"

- What are the expected analytic and informational outcomes to be produced? How will the model be used in practice? What will it predict and how will it be used? And how to update it to be better as we go.

I believe that ABC Hotels will be able to accurately predict the likelihood of a customer canceling based on historical customer data. Customers will be clustered based on the data and how it matches up with similar historical data for customers. We hope this is a model ABC Hotels can implement as new customers make reservations at their hotels.