

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.

```

> summary(corr)

```

Booking_ID	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	type_of_meal_plan	required_car_parking_space
Length:36238	Min. :0.000	Min. : 0.0000	Min. :0.0000	Min. : 0.000	Length:36238	Min. :0.00000
Class :character	1st Qu.:2.000	1st Qu.: 0.0000	1st Qu.:0.0000	1st Qu.: 1.000	Class :character	1st Qu.:0.00000
Mode :character	Median :2.000	Median : 0.0000	Median :1.0000	Median : 2.000	Mode :character	Median :0.00000
	Mean :1.845	Mean : 0.1052	Mean :0.8105	Mean : 2.204		Mean :0.03093
	3rd Qu.:2.000	3rd Qu.: 0.0000	3rd Qu.:2.0000	3rd Qu.: 3.000		3rd Qu.:0.00000
	Max. :4.000	Max. :10.0000	Max. :7.0000	Max. :17.000		Max. :1.00000
room_type_reserved	lead_time	arrival_date	market_segment_type	repeated_guest	no_of_previous_cancellations	
Length:36238	Min. : 0.00	Length:36238	Length:36238	Min. :0.00000	Min. : 0.00000	Min. : 0.00000
Class :character	1st Qu.: 17.00	Class :character	Class :character	1st Qu.:0.00000	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode :character	Median : 57.00	Mode :character	Mode :character	Median :0.00000	Median : 0.00000	Median : 0.00000
	Mean : 85.28			Mean :0.02555	Mean : 0.02335	Mean : 0.02335
	3rd Qu.:126.00			3rd Qu.:0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.00000
	Max. :443.00			Max. :1.00000	Max. :13.00000	Max. :13.00000
no_of_previous_bookings_not_canceled	avg_price_per_room	no_of_special_requests	booking_status			
Min. : 0.000	Min. : 0.00	Min. :0.00	Length:36238	Min. : 0.00000	Min. : 0.00000	Min. : 0.00000
1st Qu.: 0.000	1st Qu.: 80.30	1st Qu.:0.00		1st Qu.:0.00000	1st Qu.: 0.00000	1st Qu.: 0.00000
Median : 0.000	Median : 99.45	Median :0.00		Median :0.00000	Median : 0.00000	Median : 0.00000
Mean : 0.153	Mean :103.44	Mean :0.62		Mean :0.02555	Mean : 0.02335	Mean : 0.02335
3rd Qu.: 0.000	3rd Qu.:120.00	3rd Qu.:1.00		3rd Qu.:0.00000	3rd Qu.: 0.00000	3rd Qu.: 0.00000
Max. :58.000	Max. :540.00	Max. :5.00		Max. :1.00000	Max. :13.00000	Max. :13.00000

```

> |

```

- 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.