Executive Summary

INTRODUCTION–

In this project we have tried to understand and solve the issue of hotel booking cancellations by the customers by using the hotel_booking dataset that gives us information about the various attributes that are linked to the customer and that can be utilized to make better decisions by employing predictive modeling to solve this issue of lower occupancy at the hotel due to bookings getting cancelled.

AIM OF THE PROJECT –

The project aims to achieve a prediction model to understand and solve the issue of hotel bookings being cancelled. Our target variable, 'is_canceled' is a binary valued attribute where '0' defines 'Booking not cancelled' and '1' defines 'Booking cancelled'. The dataset provides us with a range of variables that provide interesting insights about the customer interests and the hotel bookings which we further utilize to build the model.

RESULT –

We partition the data into training-50%; validation-30%; test-20%. After performing the modeling on the dataset using the final predictor variables, we find all the models are performing good but based on the model comparison metrics we find that 2 models are performing better than other models.

Model Perfromance (On Test Data)				
Model	Accuracy of the Model	Accuarcy of 1's Prediction	AUC(Area Under Curve)	MCR
Boosted Tree	76.86%	83.04%	83.06%	23.14%
Model Average/Ensemble model	76.21%	89.98%	81.70%	23.79%
Error	0.64%	-6.94%	1.36%	-0.65%

Boosted Tree and Model average ensemble models are the final best model to compare with respect to predicting our class of interest i.e. booking is cancelled. In both modes the accuracy , Mis classification rate and Area under curve are almost silmiar but Accuracy of 1's prediction for ensemble model is 7% higher than the Boosted tree. As my class of interest is booking is cancelled which means we want to know the actual positive outcomes i.e. actual booking's cancelled from the prediction which is higher in the ensemble model. Based on these metrics we can say that Model average ensemble model is the best model which helps in solving the different business cases in the real –world industry.

APPLICATION & RECOMMENDATIONS –

After the results of prediction, we can imply that by keeping the 'lead_time' lower than 80 days in order to lower the number of cancellations. We also understand that if the hotels charge some amount at the time of booking as a convenience fee for allowing them to make changes to their booking in the future also promotes a tracked and controlled way in which the Total confirmed bookings can be estimated and Cancellations be lowered. Moreover, similarly if hotels start collecting a deposit amount to confirm the booking initially, chances are that fewer people would want to cancel or make changes to the booking now or they might lose the deposit amount. It is also recommended that the hotel charges an additional refundable amount of 25% on the bookings made by market segment 'Groups' to reduce cancellations and increase bookings.

Another way in which this prediction helps us is by providing real time updates about the availability of the rooms and amenities available on a company's website, which further also improves the traffic on the website to attract more customer base.