Strategic Mapping of Model

After the predicting threshold has been fine-tuned, we have found a balance between recall and precision. Albeit the hotel can keep increasing the recall to 100% by lowering the predicting threshold, as they can resell as many rooms as possible. In this case sacrificing too much of precision is detrimental, the hotel may overbook all of rooms which would lead to operational malfunction and even legal issues. In this section, I focused on mapping the model iteration results to the business metrics in four scenarios, for each case the following steps were executed:

1. Build a simple hotel economic model
2. Move the decision threshold and prepare corresponding confusion matrix
3. Map the confusion matrix and the economic model to find the best threshold for each case

The CatBoost model was built at five decision thresholds 0.15, 0.25, 0.4, 0.516, 0.75. From the chart, we can see True Positives Increase and True Negatives decrease as the threshold lowers. The recall metric is essential for the hotel, but there is a drastic increase on False Positives with just a slight improvement on recall.

Confusion Matrix / Decision Threshold	0.75	0.516	0.4	0.25	0.15
TP - resell, predicted canceled, and guests actually canceled	171	240	267	282	291
FN - loss, predicted show up, but guests canceled	129	60	33	18	9
FP - resell, predicted canceled, but guests showed up	35	98	147	217	273
TN - good, predicted show up, and guests actually showed up	665	602	553	483	427

The economic calculations are performed based on a hypothetical scenario for easier interpretation. Where the hotel has 1000 rooms, and they are fully booked. One room costs $200 per night, and the cost of overbooking is $50. Assuming the average show-up rate is 70% based on the dataset, the hotel immediately lists the rooms that the model predicts to be canceled for resell, at a 15% off discounted price. Now, connect revenue and losses with the confusion matrix, and find the optimal threshold for each scenario.

First case, during a normal season with 80% resell rate:

Normal Season - 80% Resell Rate	0.75	0.516	0.4	0.25	0.15
Underbook/Overbook	135	30	31	99	151
Revenue	$168,016	$185,968	$196,304	Operational Disruption	Operational Disruption
Loss	$27,040	$5,920	$1,560		
Net Profit	$140,976	$180,048	$194,744		

As we move the threshold lower, the number of True Positives increases, the hotel can list more rooms for reselling to prevent losses, thus the more revenue. The profit is maximized at threshold

0.4, if the hotel keeps increasing recall, the cost of overbooking starts to kick in, and the hotel would face operational interruptions or even legal issues, as it is against the hotel policy to overbook too many of its rooms.

Second case, during a busy season with 100% resell rate:

Busy Season - 100% Resell Rate	0.75	0.516	0.4	0.25	0.15
Underbook/Overbook	94	38	114	199	264
Revenue	$175,020	$197,460	Operational Disruption	Operational Disruption	Operational Disruption
Loss	$18,800	$1,900			
Net Profit	$156,220	$195,560			

Here, the optimal threshold is at 0.516. In a busy season, all the canceled rooms would be resold. This is when the hotel needs to be cautious and balance the precision with recall. As the threshold decreases, there is a drastic increase in False Positives. There is detrimental impact on the hotel's operational system due to too many overbookings.

Third case, during slow season with 50% resell rate:

Slow Season - 50% Resell Rate	0.75	0.516	0.4	0.25	0.15
Underbook/Overbook	197	131	93	51	18
Revenue	$157,510	$168,730	$175,190	$182,415	$187,940
Loss	$39,400	$26,200	$18,600	$10,100	$3,600
Net Profit	$118,110	$142,530	$156,590	$172,315	$184,340

During a low-demand season, the hotel has the highest profits at 0.15 threshold. Even though, the model makes more mistakes by classifying non-canceled bookings as canceled. Due to low resell rate, the hotel is less prone to overbook. Therefore, lower the threshold, means higher recall, which leads to less empty rooms, and the hotel is minimizing its losses.

Conclusion

In summary, the model had successfully saved hotels from large portion of financial loss due to cancellations. In light of optimizing model, the hotel should exercise caution in picking the most suitable decision threshold. My suggestion to the hotel is to implement dynamic thresholding based on business priorities and booking patterns. For example, during times of high demand or when rooms are nearly sold out, the hotel should focus on precision (to avoid overbookings), while during low-demand seasons, the hotel may prioritize recall (to avoid missed revenue opportunities).