

Hotel Cancellation Prediction



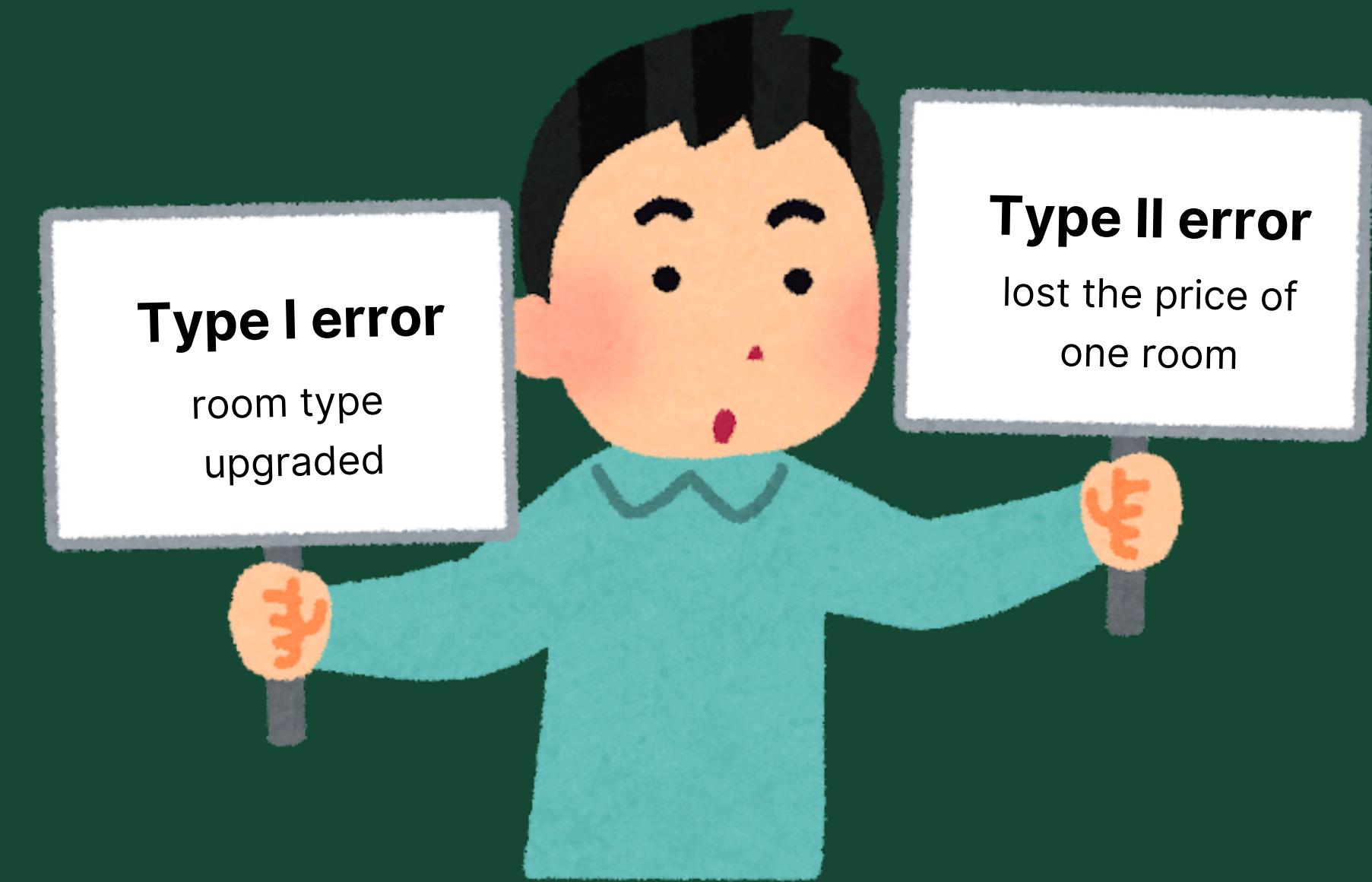
Problem:

The hospitality industry is currently grappling with a significant challenge:

Booking Cancellations.

source : <https://bettertravel.ai/hotel-prices/portugal>
<https://www.americanexpress.com/in/articles/life-with-amex/travel/hotel-room-upgrade>

Target :



Metric Evaluation : F2 (emphasis on recall to reduce Type II errors) will be used to evaluate the model's performance.

Consequences

Type I Error

Average loss per false positive: \$12 to \$59



Decreased Guest Loyalty

First-time guests are also less likely to rebook if they are disappointed by overbooking.



Negative Brand Impact

Word of mouth is a powerful marketing tool. If things go wrong, your brand image can be damaged and it can take a long time to recover.

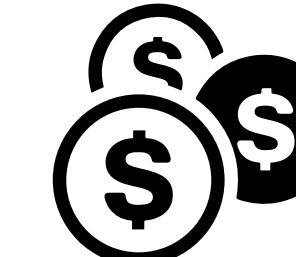


Financial Loss

Hotels will offer compensation in the form of additional amenities, bonuses, and room upgrades.

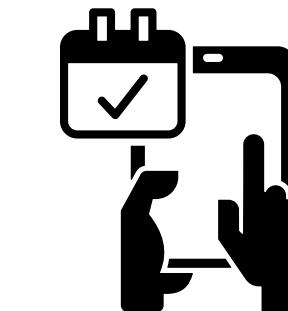
Type II Error

Revenue loss per false negative: \$110 per night



Lower Revenue

Any room that's vacant in a hotel is probably a loss-making entity for the day.



Ranking Drop on OTA

While OTAs offer free cancellations, they can affect your rankings, sales, and revenue.



Planning Disruptions

Hotels must plan for the number of guests they expect. If guests don't show up, there might be a shortage of resources.

DATA UNDERSTANDING

Data
Overview

11 columns
83573 rows

Potential
Problem



missing value in column 'country'
outliers in numerical column
80174 duplicated data
high cardinality in column 'country'
data imbalance

Feature
Engineering

One Hot Encoding
Binary Encoding

Experiment Design

EXPERIMENT 1

Focuses on basic setup
(categorical encodings only)

EXPERIMENT 2

Basic setup + Robust scaling

EXPERIMENT 3

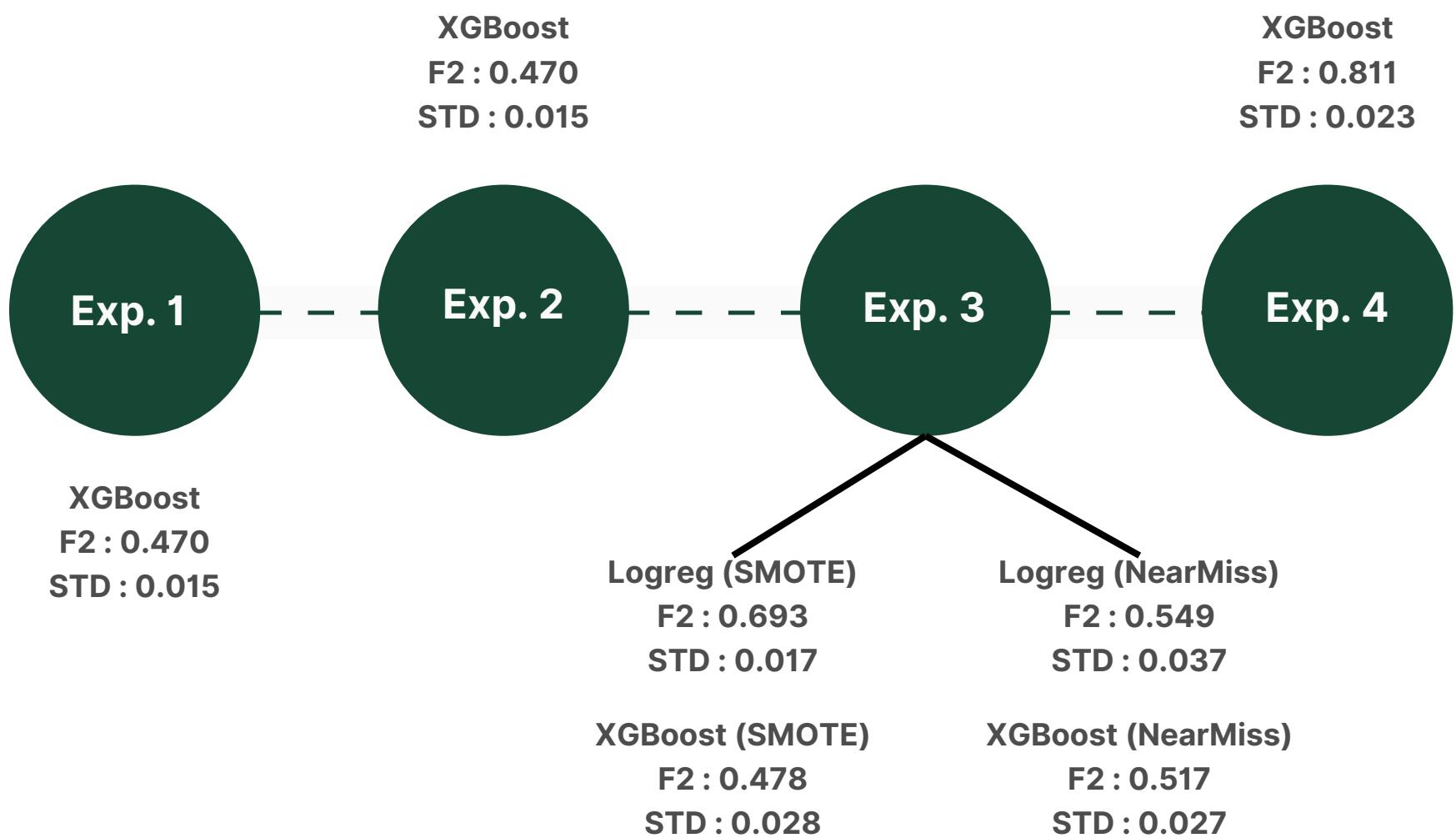
Basic setup + Resampling
(SMOTE, NearMiss)

EXPERIMENT 4

Scaling + Resampling +
Tuning



The Best Performing Model

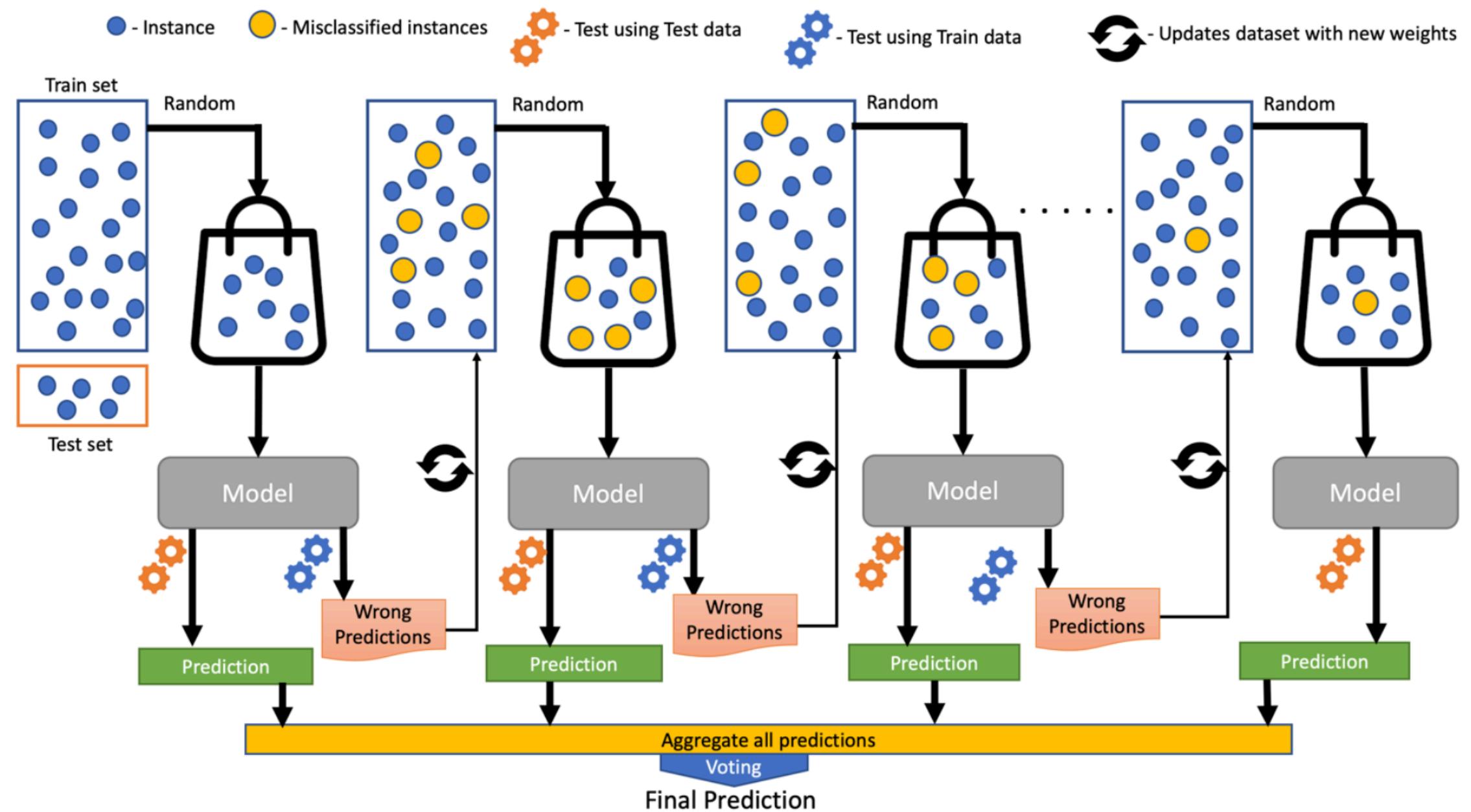


Insight:

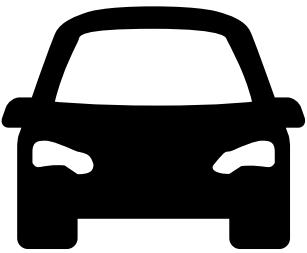
XGBoost appears to be a strong baseline across all experiments. It performs well (F2 score above 0.47) in 3 experiments with the best performance at 0.811.

	Train Set	Test Set
Accuracy	70,87%	66.43%
Precision	51.33%	47.15%
Recall	94.92%	90.75%
F2 score	81.14%	76.59%

Model's explanation



Customer Cancellation Factors!



required_car_parking_space

Without parking, customers cancel 39.3% of the time. This shows that parking affects customer cancellations.



deposit_type

Non-refundable deposits don't stop customers from cancelling (99.4%)



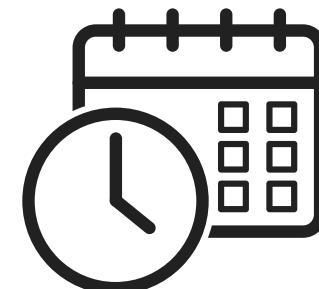
market_segment

61% of customers in the market segment group cancel. Over half of hotel bookings in this market segment were cancelled.



booking_changes

Cancelled orders have fewer average changes than undelivered orders.

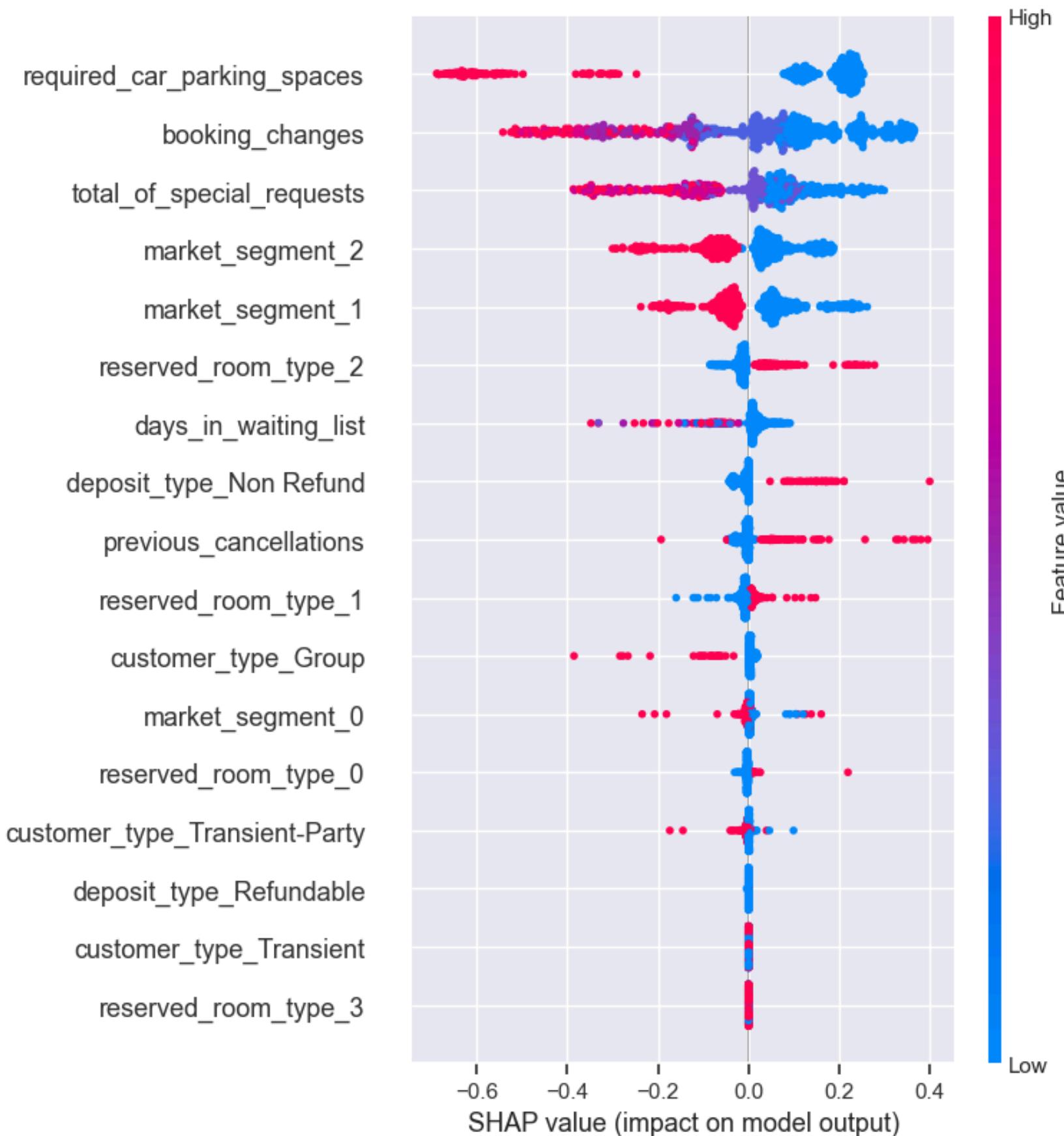


days_in_waiting_list

The longer customers have to wait for their order to be confirmed, the more likely they are to cancel.

SHAP

SHAP value: This shows how each feature affects the model's predictions. A positive value means the feature increases the probability of cancellation, while a negative value means it decreases it. The color indicates the value of the feature. Red has a high value, while blue has a low value.



Implementation

Without Model

**Assume 60% incorrect predictions with
FP of 20% and FN 40%**

FP: $0.2 \times 572 \text{ bookings} \times \$59 = \$6,749.6$

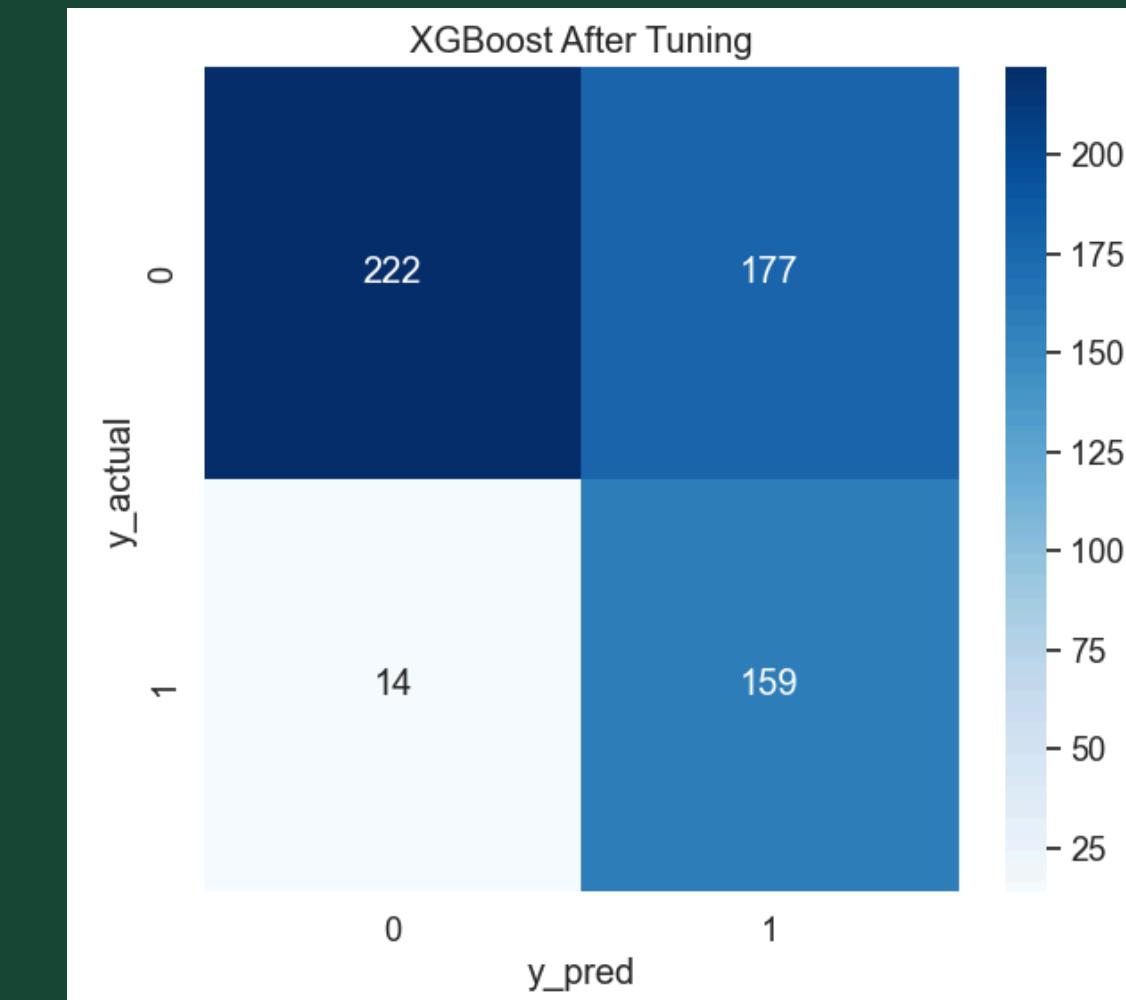
FN: $0.4 \times 572 \text{ bookings} \times \$110 = \$25,168$

Total Financial Lost : \$31,917.2



\$19,934.6
62%
saving with model

With Model

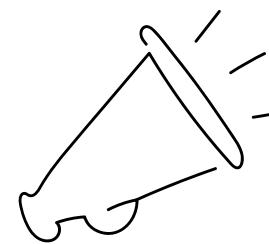


- FP (Type I error): actual 0, pred 1
 $177 \times \$59 = \$10,443$
- FN (Type II error): actual 1, pred 0
 $14 \times \$110 = \$1,540$
- Total Financial Lost: \$11,983

Conclusion

XGBoost model in Experiment 4 successfully balanced precision and recall, significantly enhancing the prediction of hotel booking cancellations. By using this model, hotels can proactively manage room inventory and reduce potential financial losses due to sudden cancellations.

Recommendations



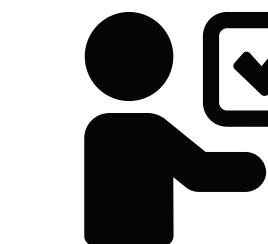
Highlight parking availability in booking confirmations and marketing materials to reassure customers.



Revisit the non-refundable deposit policy.



Identify market segments that have high cancellation rates and develop targeted strategies



Speed up the booking confirmation process to reduce waiting time.

**Let's viewing prediction results in
Google Cloud Platform!**