

DATA MINING PRESENTATION

Hawk Hotel Consultants



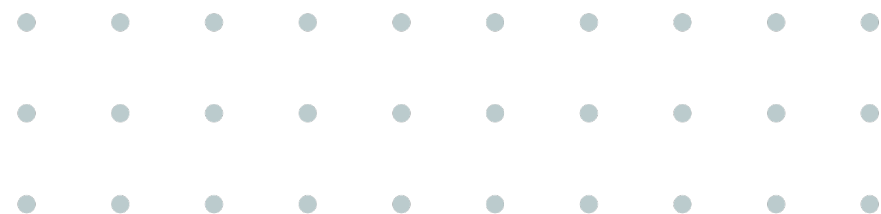
01. BACKGROUND INFORMATION

02. OUR GOALS

03. SOLUTION & RECOMMENDATIONS

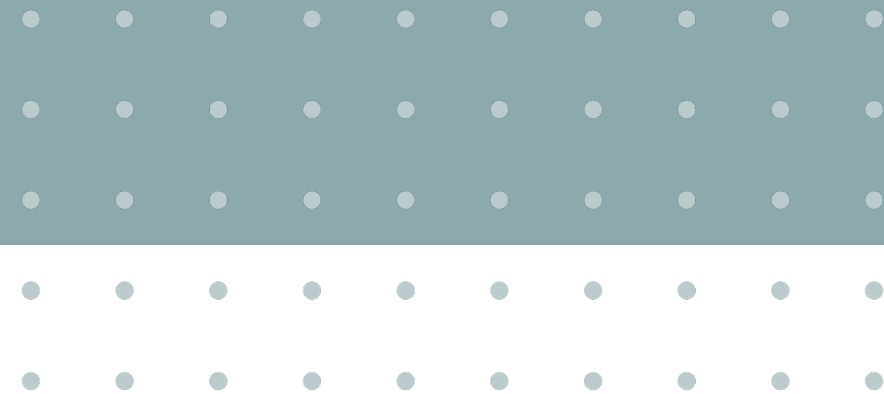
04. LIMITATIONS

AGENDA




01.

BACKGROUND INFORMATION



PROBLEM DESCRIPTION



**33% of rooms
were cancelled**
in 2017 & 2018

- Industry average was 12.8% in 2018
- Find strategies to keep hotel fully booked
- Target high cancellations & build customer loyalty



BACKGROUND
INFORMATION

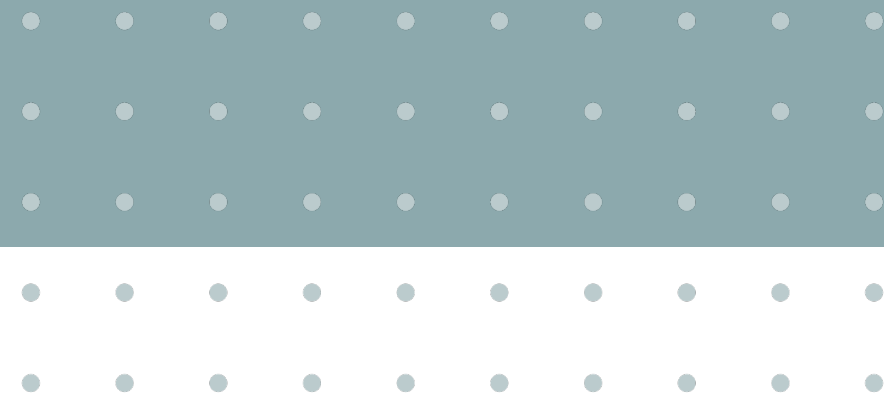
OUR
GOAL

SOLUTION

LIMITATIONS

02.

OUR GOALS



BUSINESS GOAL

- Keep hotel rooms fully booked to maximize profits
- Maintain customer satisfaction



BUSINESS GOAL

- Keep hotel rooms fully booked to maximize profits
- Maintain customer satisfaction

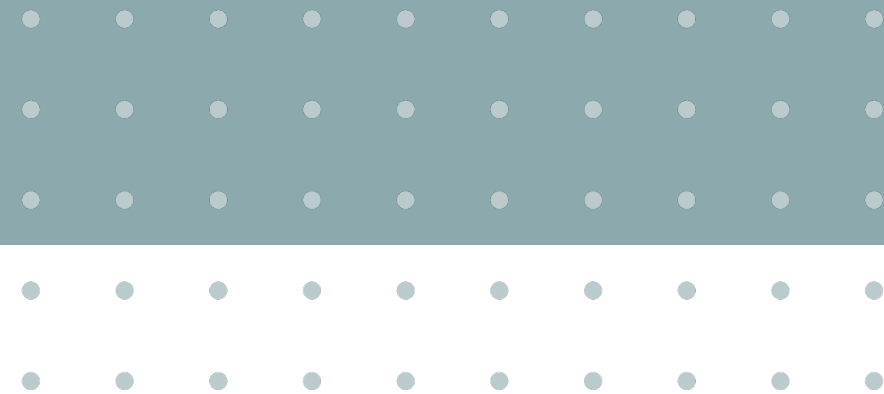
DATA MINING GOAL

- Determine most important factors that lead to cancellations
- Build model that predicts cancellations



03.

SOLUTIONS & RECOMMENDATIONS



SOLUTION #1

Training

Model	AUC	CA	F1	Precision	Recall	MCC
Neural Network	0.885	0.825	0.821	0.822	0.825	0.593
Decision Tree	0.943	0.877	0.876	0.875	0.877	0.718
Random Forest	0.911	0.857	0.853	0.856	0.857	0.668
Gradient Boosting	0.893	0.836	0.826	0.840	0.836	0.616

Testing

Model	AUC	CA	F1	Precision	Recall	MCC
Neural Network	0.878	0.819	0.815	0.815	0.819	0.573
Decision Tree	0.916	0.851	0.850	0.850	0.851	0.655
Random Forest	0.903	0.855	0.851	0.853	0.855	0.658
Gradient Boosting	0.888	0.833	0.823	0.835	0.833	0.602

Cross Validation (10 Fold)

Model	AUC	CA	F1	Precision	Recall	MCC
Neural Network	0.883	0.822	0.819	0.819	0.822	0.587
Decision Tree	0.920	0.858	0.857	0.856	0.858	0.675
Random Forest	0.906	0.852	0.847	0.852	0.852	0.656
Gradient Boosting	0.890	0.834	0.824	0.838	0.834	0.612

- Predictive model – classifies as “cancel” or “not cancel”
- Decision tree provides the highest AUC and does not overfit

BACKGROUND
INFORMATION

OUR
GOAL

SOLUTION

LIMITATIONS

SOLUTION #1

Decision Tree Model Factors

Number of adults

Number of children

Number of weekend nights

Number of week nights

Meal plan

Parking space

Room type reserved

Lead time

Arrival month

Market segment type

Repeat guest

of previous cancellations

of bookings not canceled

Average price per room

Special requests

BACKGROUND
INFORMATION

OUR
GOAL

SOLUTION

LIMITATIONS

SOLUTION #1

Decision Tree Model Factors

Number of adults
Number of children
Number of weekend nights
Number of week nights
Meal plan
Parking space
Room type reserved
Lead time

Arrival month
Market segment type
Repeat guest
of previous cancellations
of bookings not canceled
Average price per room
Special requests

Result

Likely to cancel

Unlikely to cancel

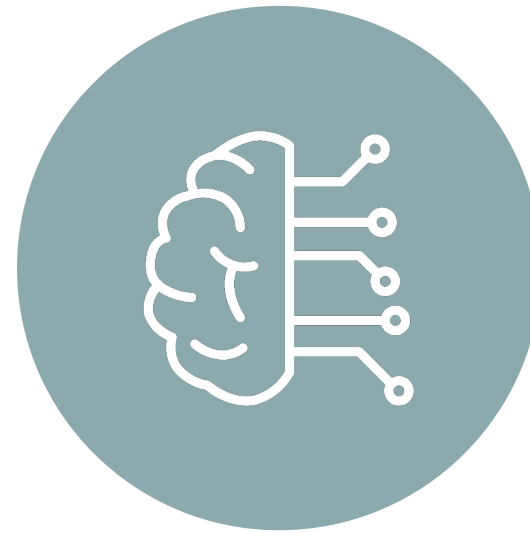
BACKGROUND
INFORMATION

OUR
GOAL

SOLUTION

LIMITATIONS

RECOMMENDATION #1



Implement this predictive model onto reservation website to determine the probability of reservation cancellation.

BACKGROUND
INFORMATION

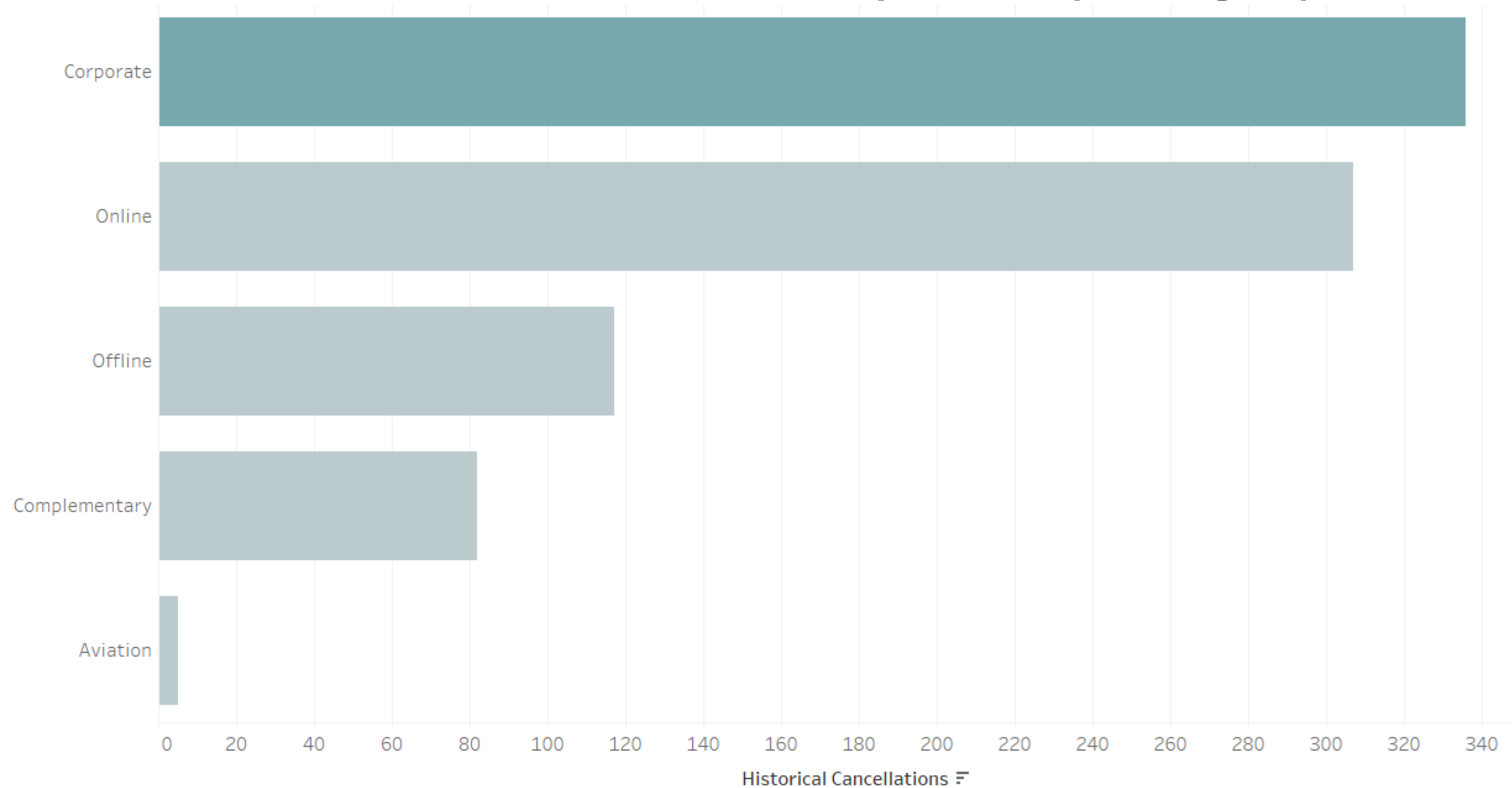
OUR
GOAL

SOLUTION

LIMITATIONS

SOLUTION #2

Cancellations have come mostly from corporate groups.



BACKGROUND
INFORMATION

OUR
GOAL

SOLUTION

LIMITATIONS

RECOMMENDATION #2



Target corporate groups by offering incentives and personalized services, or raise cancellation fees to decrease chances of cancellation.

BACKGROUND
INFORMATION

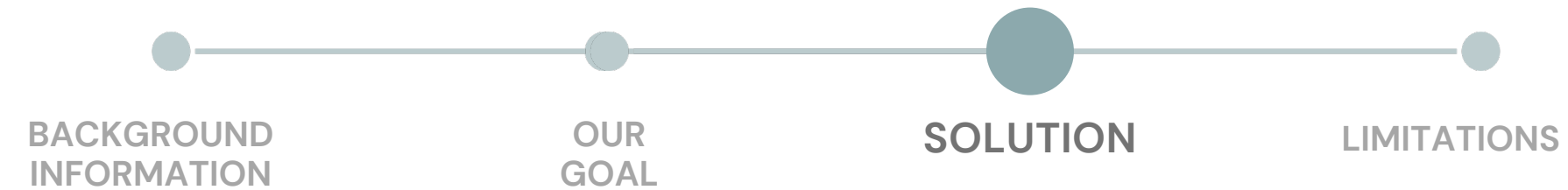
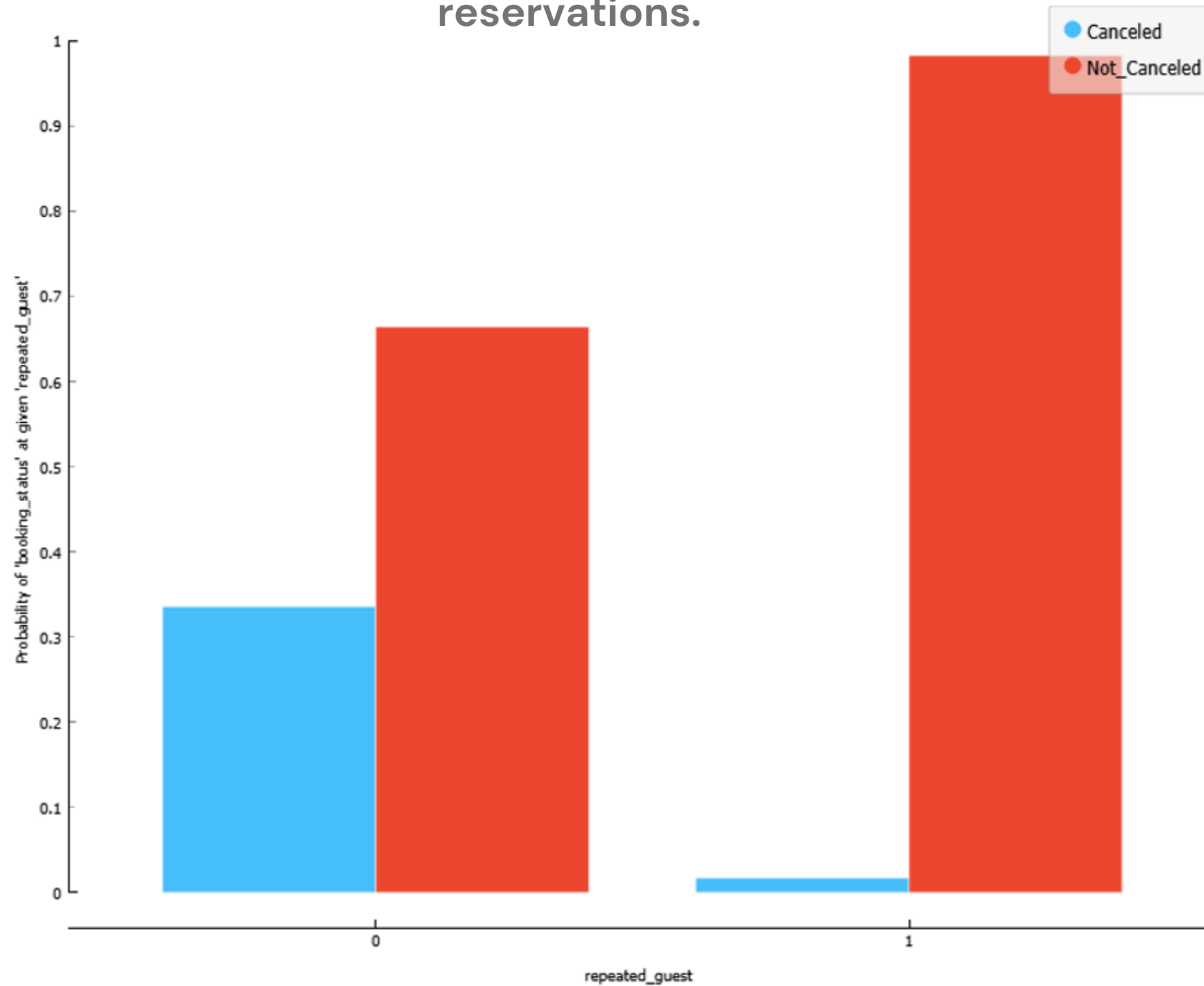
OUR
GOAL

SOLUTION

LIMITATIONS

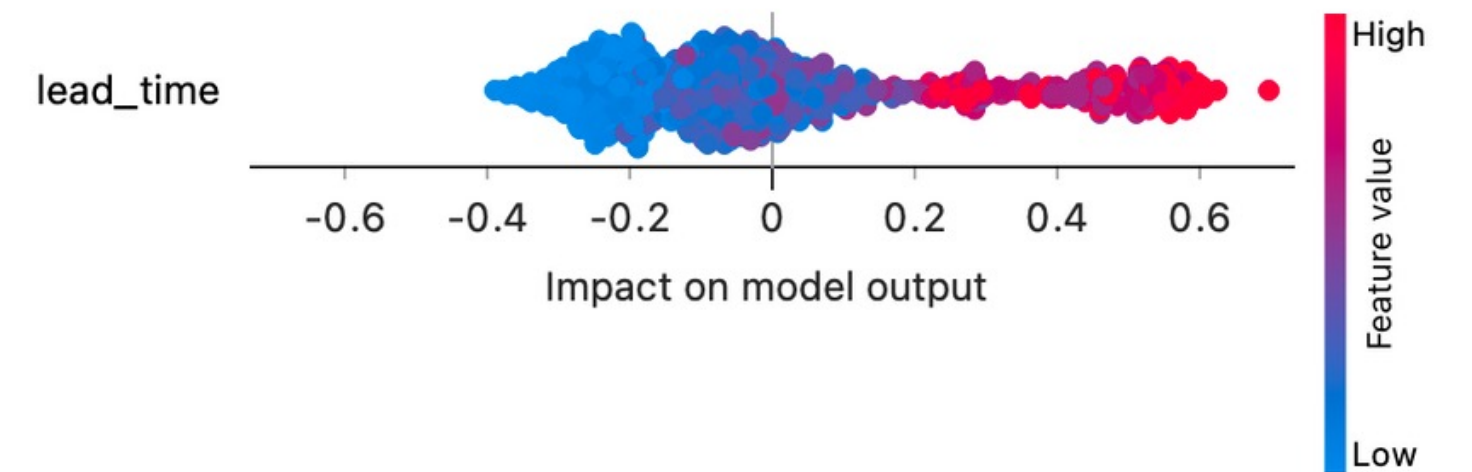
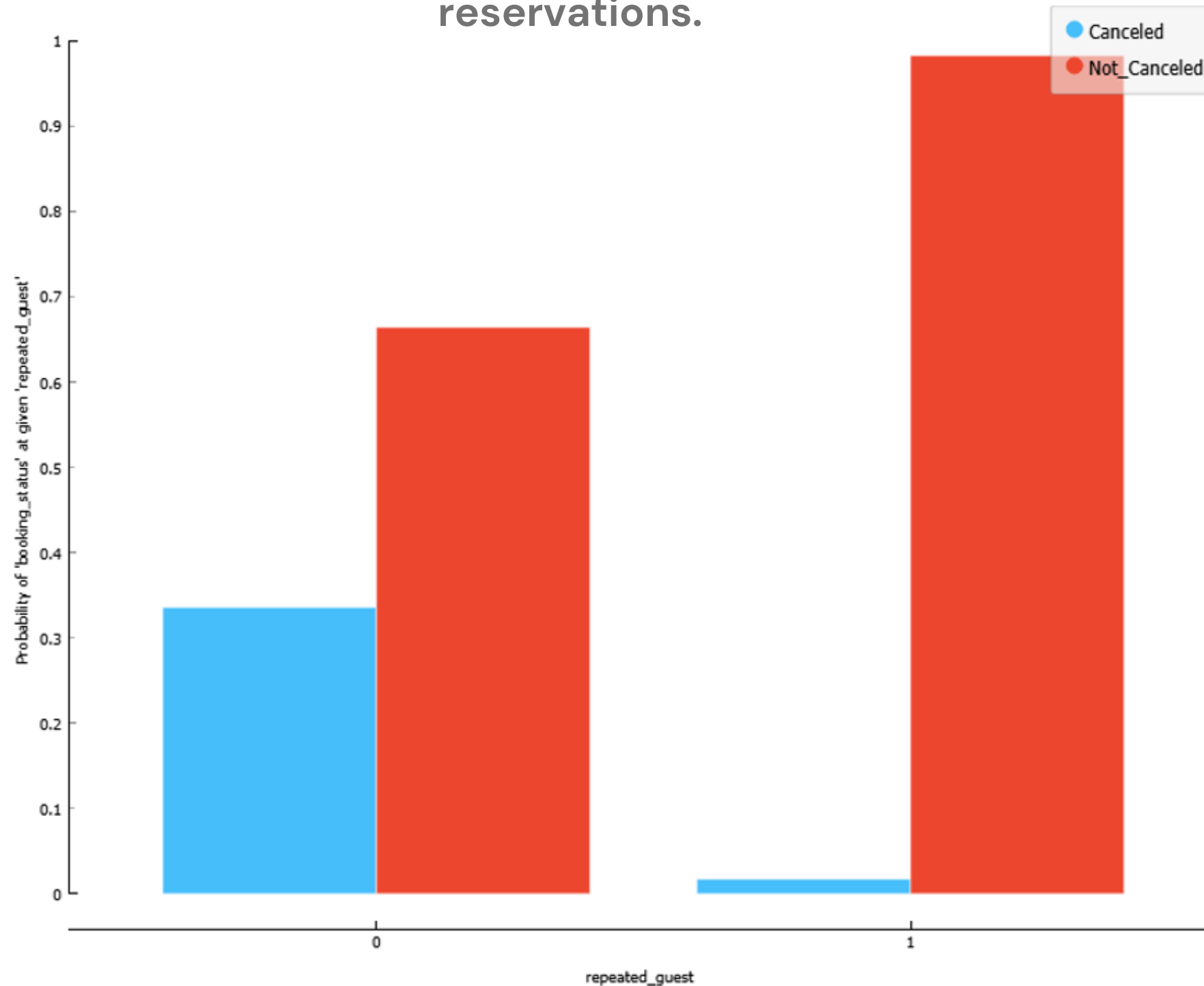
SOLUTION #3

Repeat guests are less likely to cancel their hotel reservations.



SOLUTION #3

Repeat guests are less likely to cancel their hotel reservations.



Lead time has the greatest impact on guest cancellations.



RECOMMENDATION #3



Create a loyalty program that prioritizes repeat guests in the booking process, allowing them to reserve rooms earlier.

BACKGROUND
INFORMATION

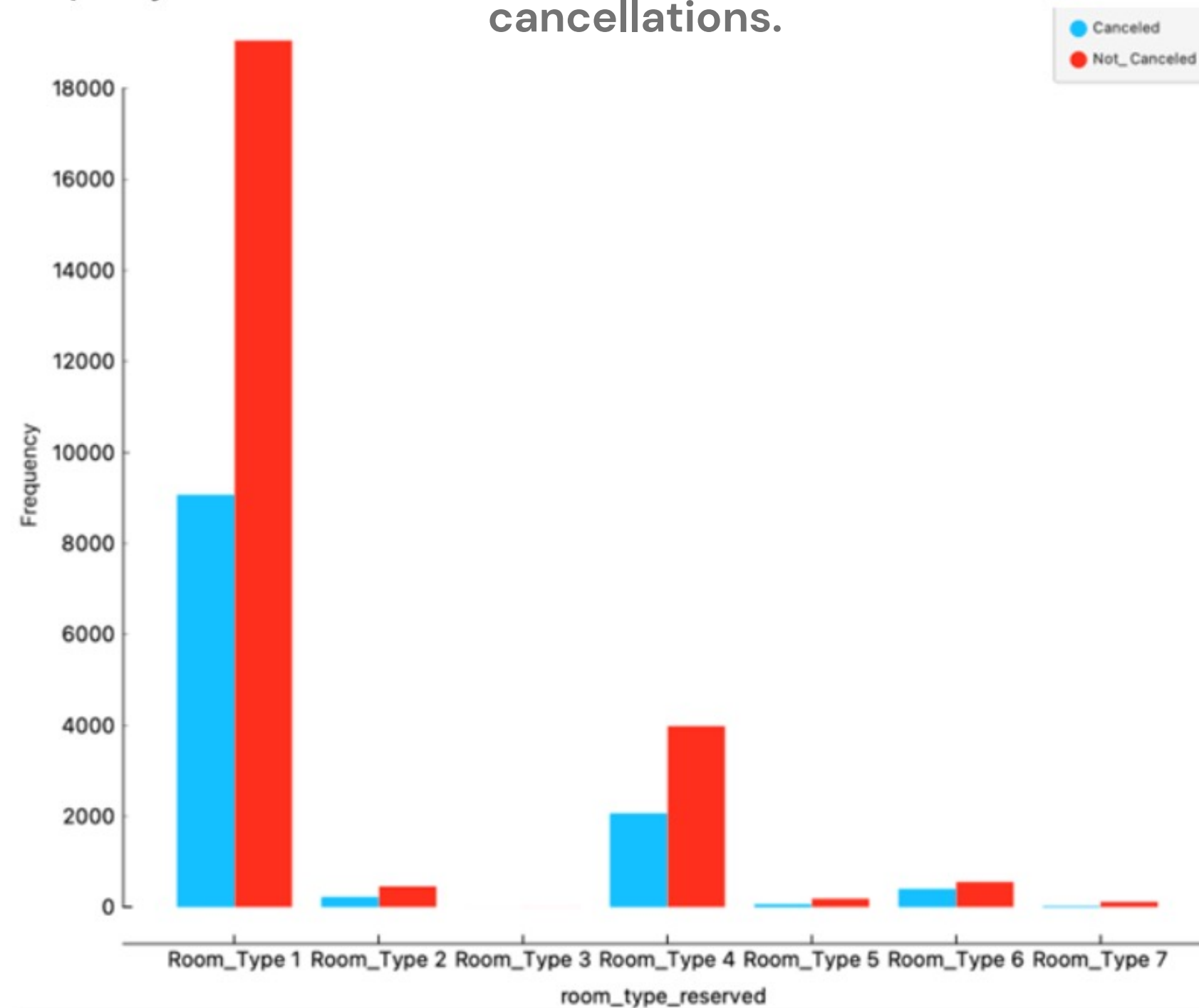
OUR
GOAL

SOLUTION

LIMITATIONS

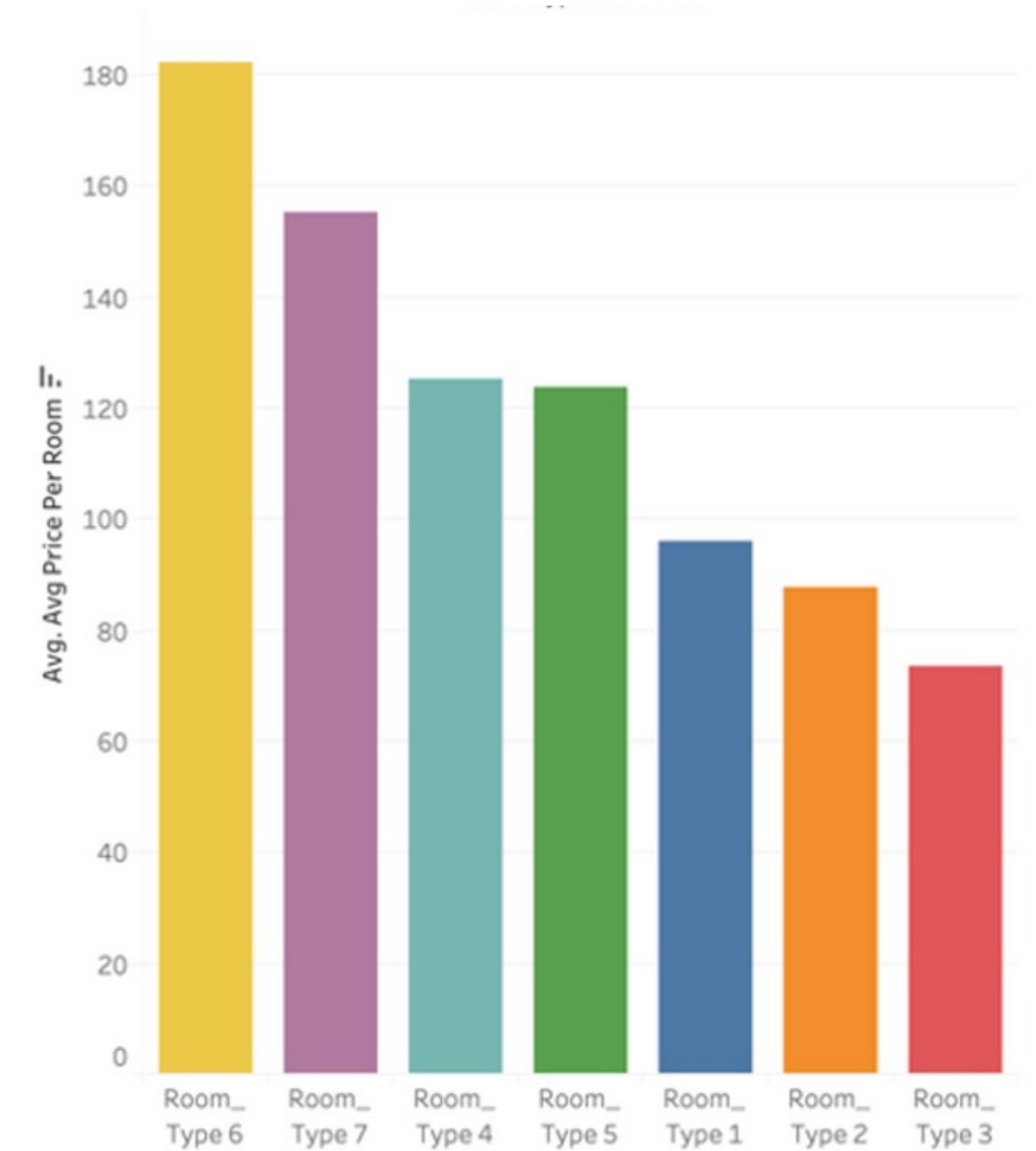
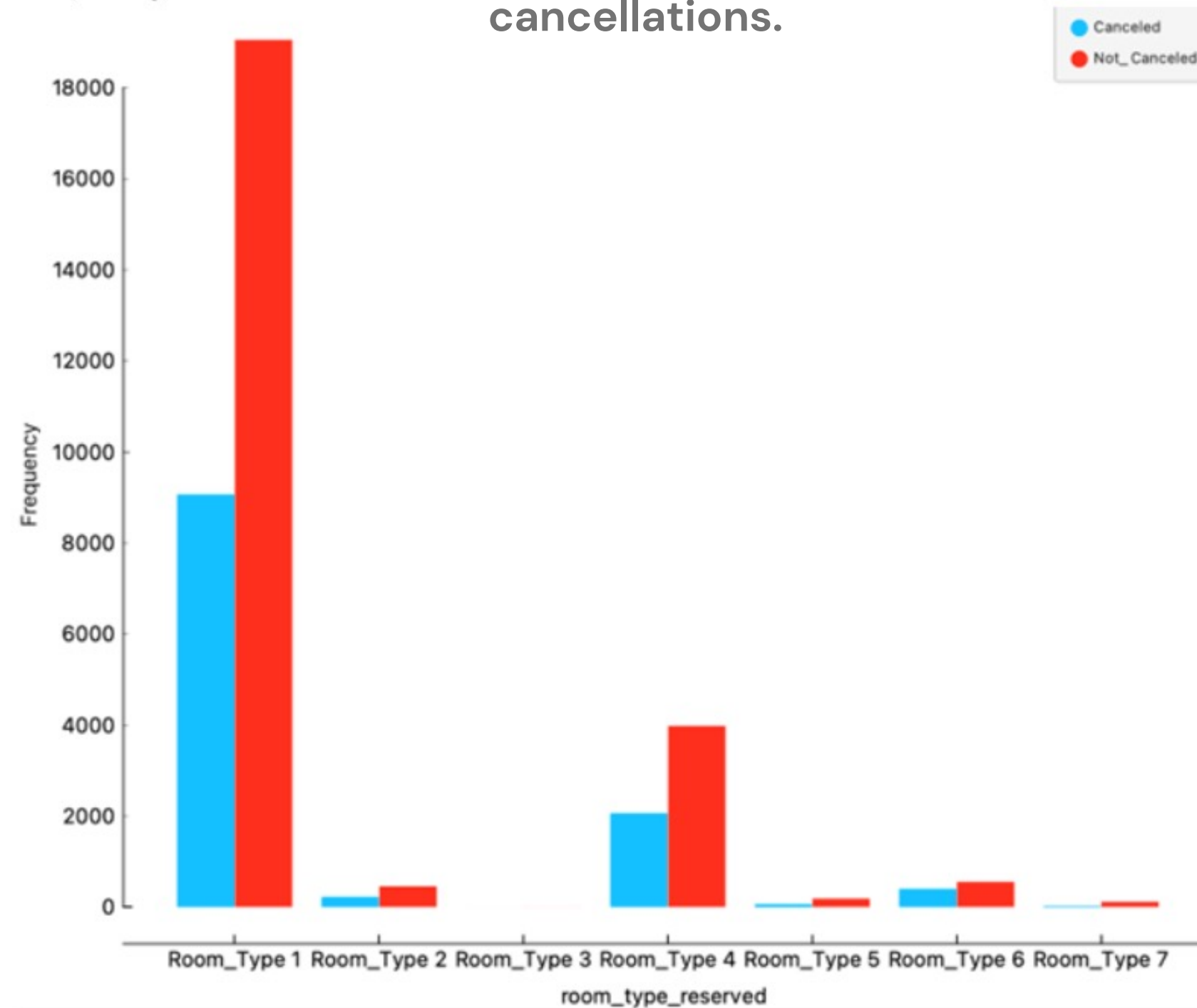
SOLUTION #4

Room Type 1 has the most bookings and the most cancellations.



SOLUTION #4

Room Type 1 has the most bookings and the most cancellations.



Room Types 1 and 4 are moderately priced, relative to other rooms.



RECOMMENDATION #4



Restructure what Room Type 1 and Room Type 4 offer guests and how the value proposition is communicated to customers.

BACKGROUND
INFORMATION

OUR
GOAL

SOLUTION

LIMITATIONS

04.

LIMITATIONS



LIMITATIONS

Number of beds,
bathroom
configurations, and
other amenities not
provided in dataset

Pre-COVID Data
(2017–2018)

Cancellations are
circumstantial

BACKGROUND
INFORMATION

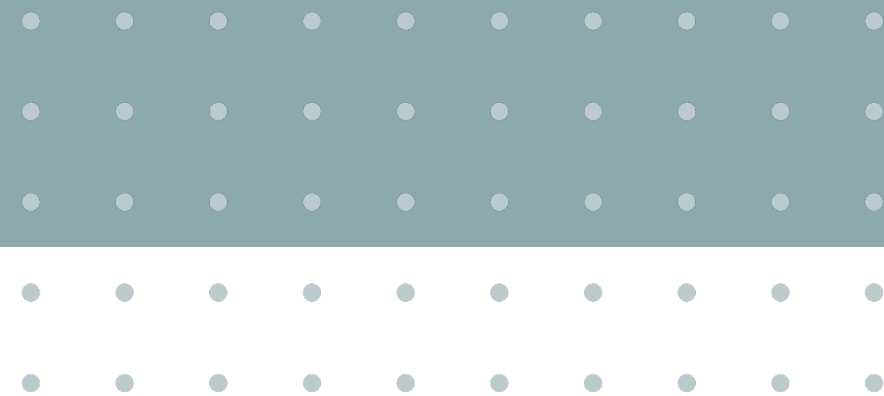
OUR
GOAL

SOLUTION

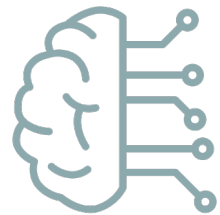
LIMITATIONS

05.

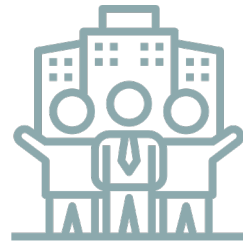
CONCLUSION



CONCLUSION



Implement predictive model on website to determine probability of reservation cancellation.



Target corporate groups by offering incentives and raise cancellation fees to decrease chances of cancellation.



Create a loyalty program that prioritizes repeat guests, allowing them to reserve rooms earlier.



Restructure what Room Type 1 and Type 4 offer and how the value proposition is communicated to guests.

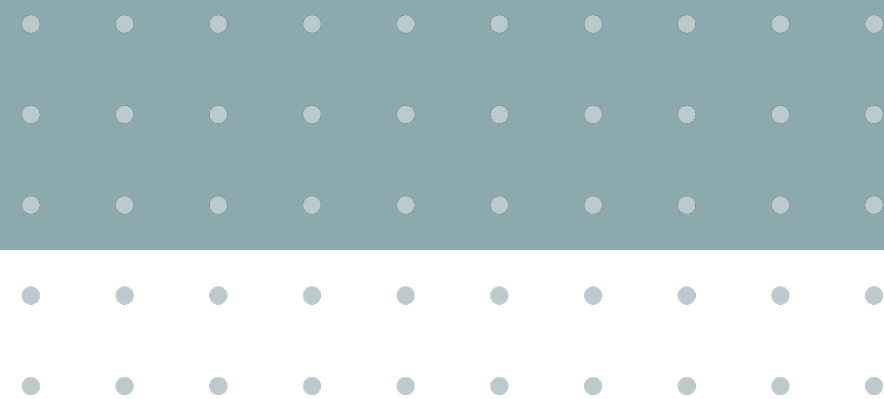
06.

QUESTIONS?



07.

WORKS CITED



CITATIONS

Raza, A. (2023, January 4). Hotel Reservations Dataset. Kaggle.

<https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>

Revenue Hub. (2020). Cancellations Rates: Where do They Stand and How to Overcome Them?

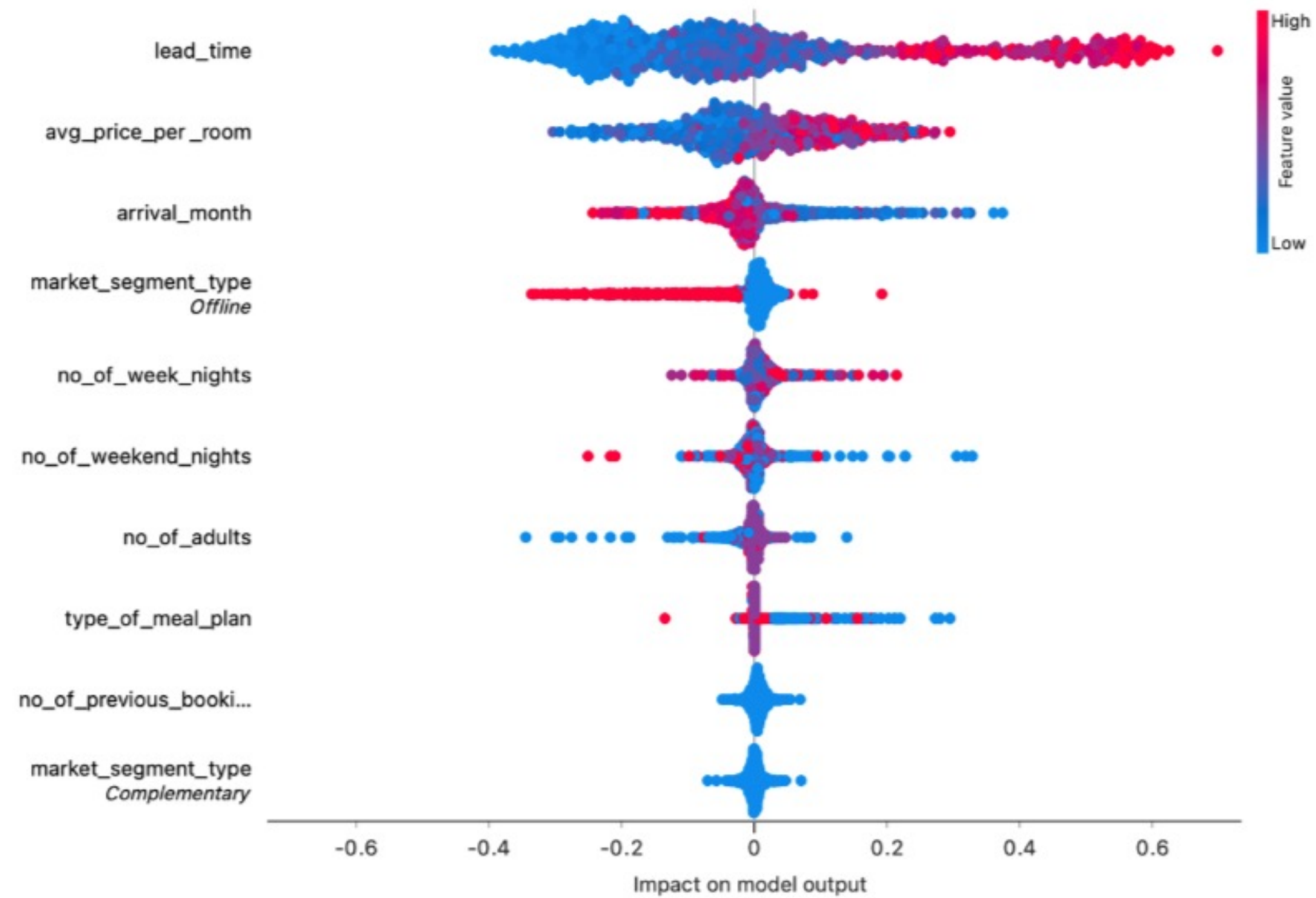
<https://revenue-hub.com/cancellations-rates-where-do-they-stand-and-how-to-overcome-them/>

08.

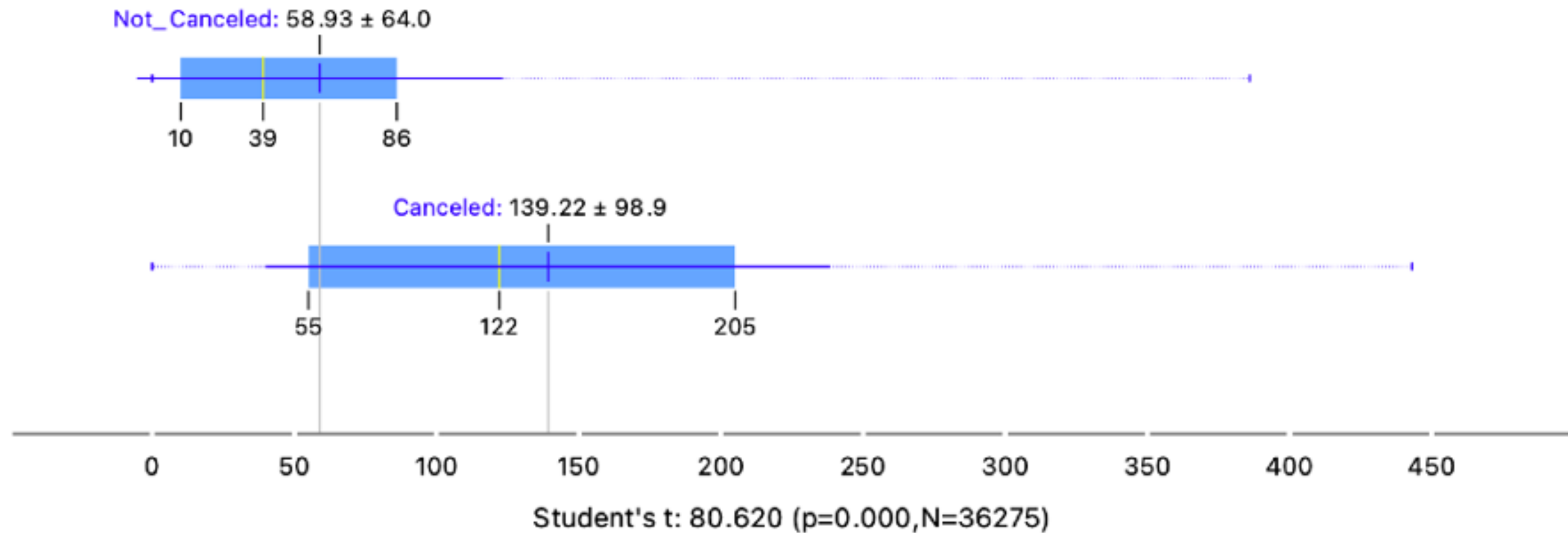
APPENDIX



Lead time, price, and arrival month have the largest impact on whether a guest will cancel their room.



Hotel reservations booked closer to a guest's arrival date are less likely to be canceled.



Box plot for attribute 'lead_time' grouped by 'booking_status'

Decision Tree – Confusion
Matrix

Predicted

		Canceled	Not_Canceled	Σ
Actual	Canceled	6598	1782	8380
	Not_Canceled	1351	15662	17013
Σ		7949	17444	25393

Parameters for Models

Gradient Boosting - Ora... ? X

Name
Gradient Boosting

Method
Gradient Boosting (scikit-learn) v

Basic Properties

Number of trees: 108

Learning rate: 0.011

☒ Replicable training

Growth Control

Limit depth of individual trees: 4

Do not split subsets smaller than: 4

Subsampling

Fraction of training instances: 1.00

☒ Apply Automatically

≡ ? | 25.4k | - | □ | M

Random Forest - Orange ? X

Name
Random Forest

Basic Properties

Number of trees: 90

☒ Number of attributes considered at each split: 10

☒ Replicable training

☐ Balance class distribution

Growth Control

☒ Limit depth of individual trees: 6

☒ Do not split subsets smaller than: 6

☒ Apply Automatically

≡ ? | 25.4k | - | □ | M

4,4 - Orange ? X

Name
4,4

Neurons in hidden layers: 4,4

Activation: ReLu v

Solver: Adam v

Regularization, $\alpha=0.01$:

Maximal number of iterations: 100

☒ Replicable training

Cancel ☒ Apply Automatically

≡ ? | 25.4k | - | □ | M

Tree - Orange ? X

Name
Tree

Parameters

☒ Induce binary tree

☒ Min. number of instances in leaves: 30

☒ Do not split subsets smaller than: 20

☒ Limit the maximal tree depth to: 200

Classification

☒ Stop when majority reaches [%]: 95

☒ Apply Automatically

≡ ? | 25.4k | - | □ | M