HOTEL CANCELLATIONS

PROBLEM

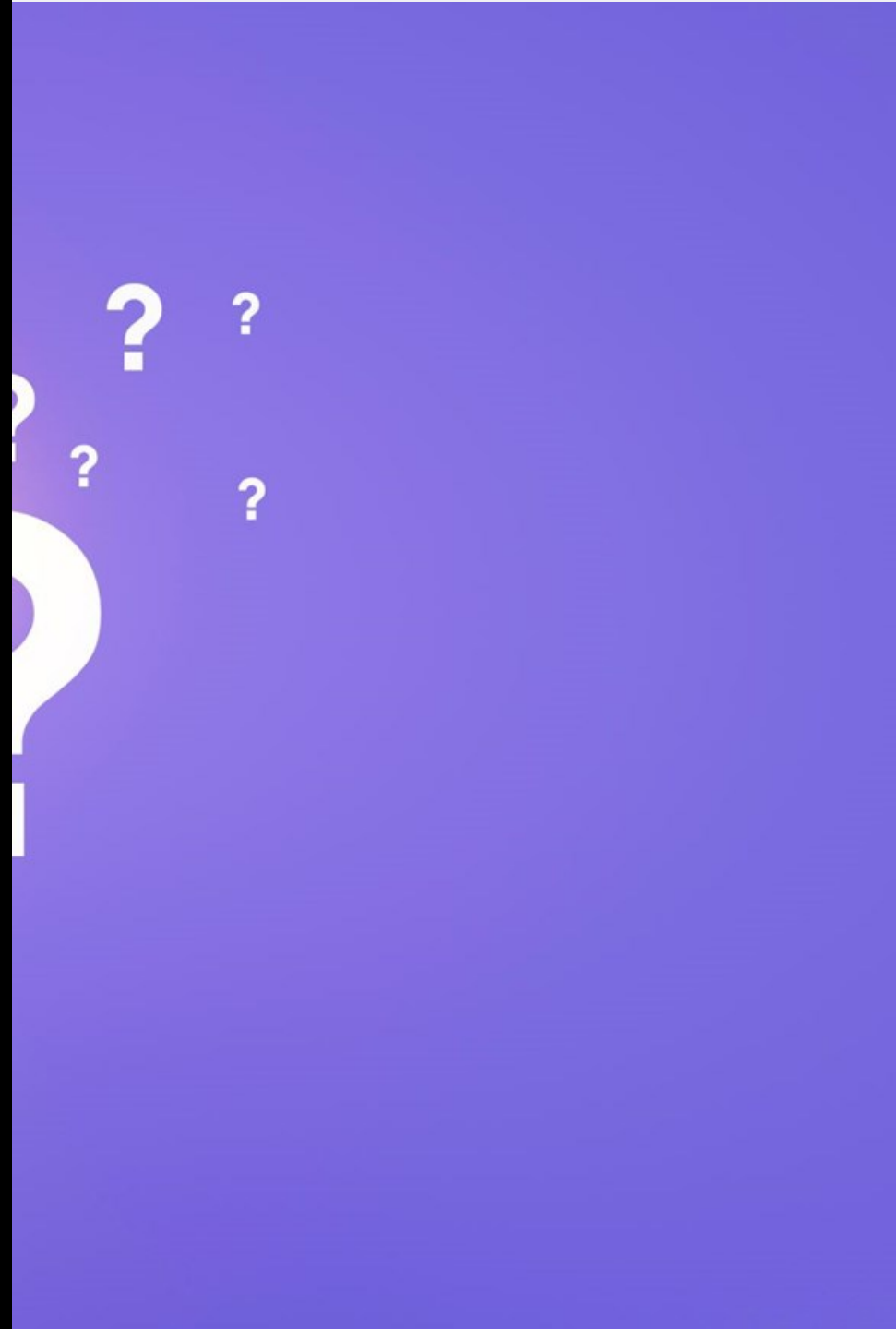
TACKLE THE ISSUE OF HOTEL BOOKING CANCELLATIONS

WHY?

UNDERSTAND HOW A CANCELLATION CAN BE AVOIDED

HOW?

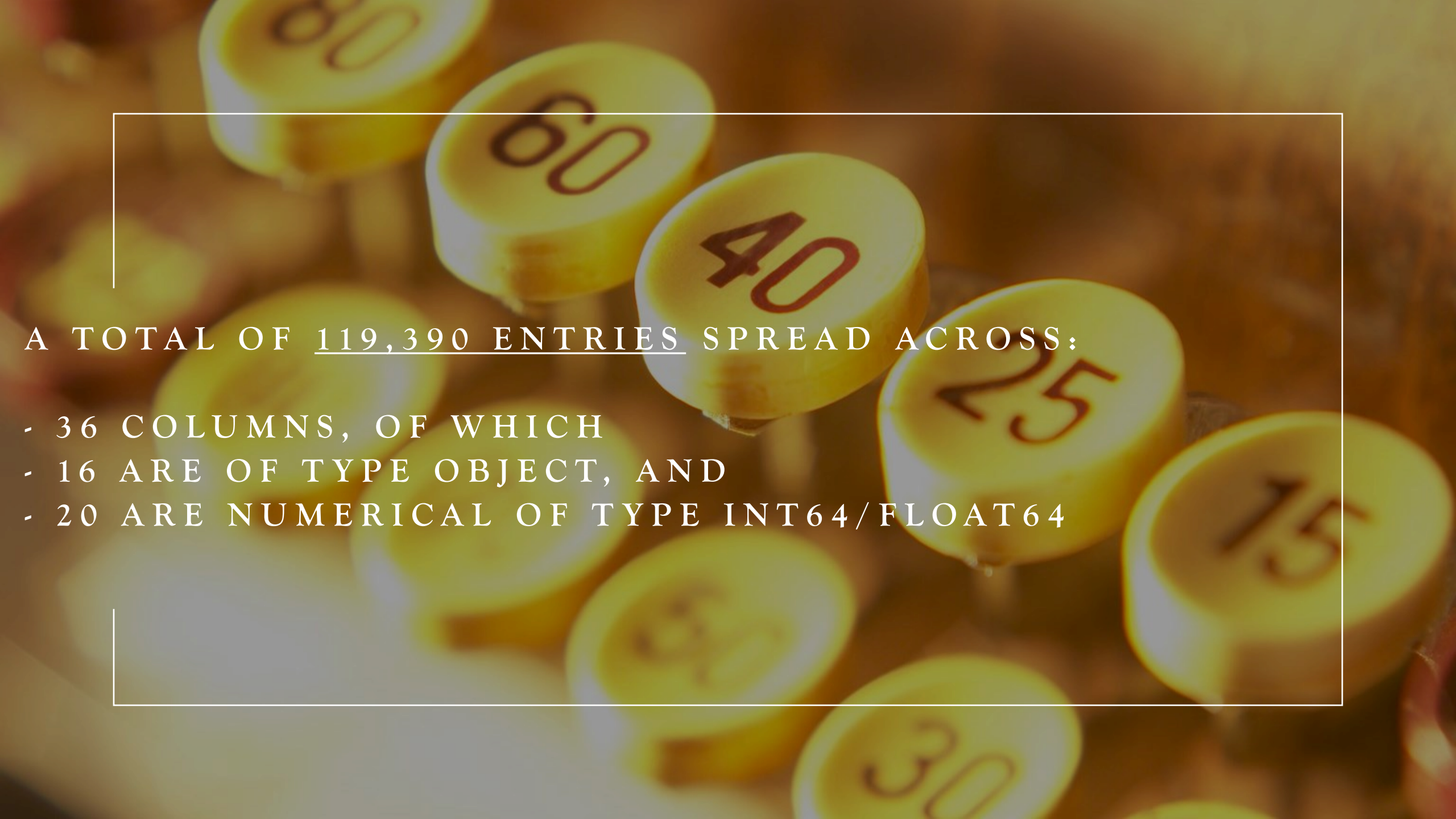
BY BUILDING A MODEL TO PREDICT WHETHER A RESERVATION WILL BE CANCELED (1) OR NOT (0)



SUPERVISED MODEL

DESCRIPTION OF THE DATASET

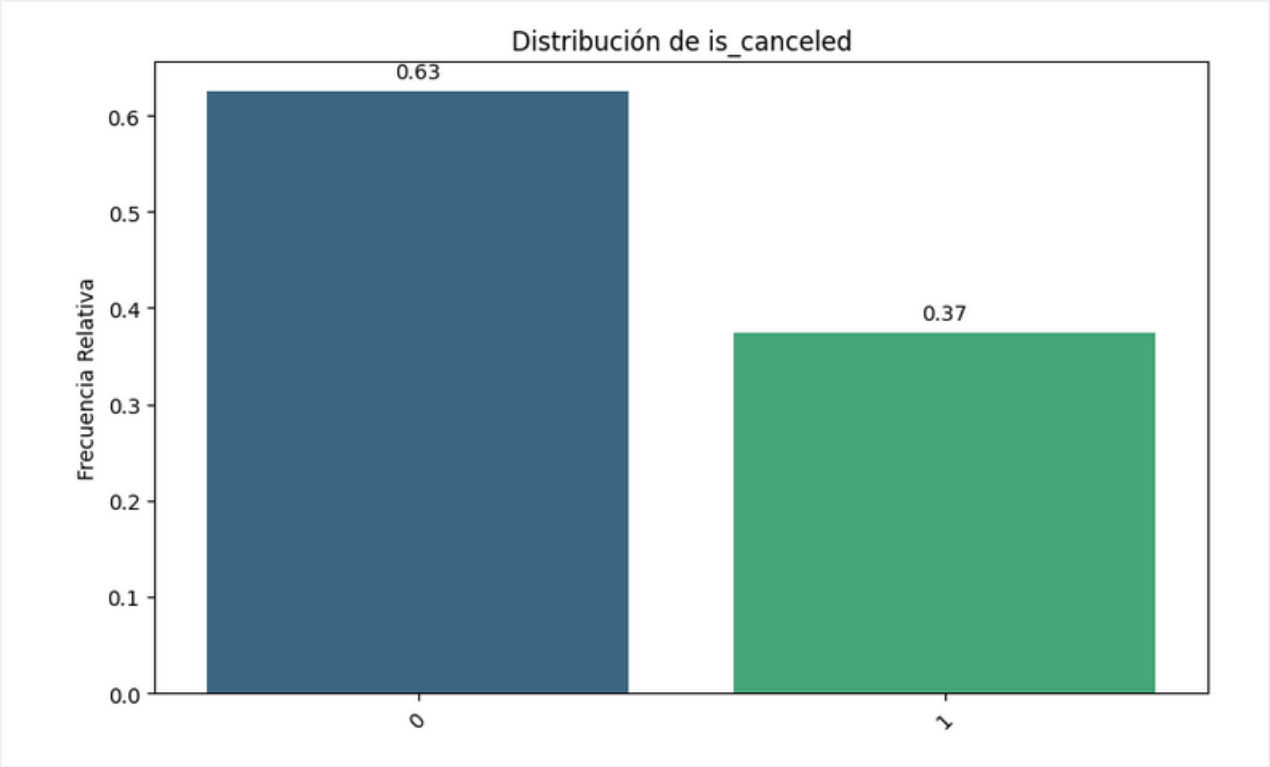
Index	Variable	Description
1	hotel	Type of hotel (Resort Hotel, City Hotel)
2	**is_canceled**	**Reservation cancellation status (0 = not canceled, 1 = canceled)**
3	lead_time	Number of days between booking and arrival
4	arrival_date_year	Year of arrival
5	arrival_date_month	Month of arrival
6	arrival_date_week_number	Week number of the year for arrival
7	arrival_date_day_of_month	Day of the month of arrival
8	stays_in_weekend_nights	Number of weekend nights (Saturday and Sunday) the guest stayed or booked
9	stays_in_week_nights	Number of week nights the guest stayed or booked
10	adults	Number of adults
11	children	Number of children
12	babies	Number of babies
13	meal	Type of meal booked (BB, FB, HB, SC, Undefined)
14	country	Country of origin of the guest
15	market_segment	Market segment designation
16	distribution_channel	Booking distribution channel
17	is_repeated_guest	If the guest is a repeat customer (0 = not repeated, 1 = repeated)
18	previous_cancellations	Number of previous bookings that were canceled by the customer
19	previous_bookings_not_canceled	Number of previous bookings that were not canceled by the customer
20	reserved_room_type	Type of reserved room
21	assigned_room_type	Type of assigned room
22	booking_changes	Number of changes made to the booking
23	deposit_type	Type of deposit made (No Deposit, Refundable, Non Refund)
24	agent	ID of the travel agent responsible for the booking
25	company	ID of the company responsible for the booking
26	days_in_waiting_list	Number of days the booking was in the waiting list
27	customer_type	Type of customer (Transient, Contract, Transient-Party, Group)
28	adr	Average Daily Rate
29	required_car_parking_spaces	Number of car parking spaces required
30	total_of_special_requests	Number of special requests made
31	reservation_status	Last reservation status (Check-Out, Canceled, No-Show)
32	reservation_status_date	Date of the last reservation status
33	name	Guest's name
34	email	Guest's email address
35	phone-number	Guest's phone number
36	credit_card	Last four digits of the guest's credit card



A TOTAL OF 119,390 ENTRIES SPREAD ACROSS:

- 36 COLUMNS, OF WHICH
- 16 ARE OF TYPE OBJECT, AND
- 20 ARE NUMERICAL OF TYPE INT64/FLOAT64

DISTRIBUTION OF THE TARGET


Distribución de is_canceled
Frecuencia Relativa
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.63
0.37
0
1

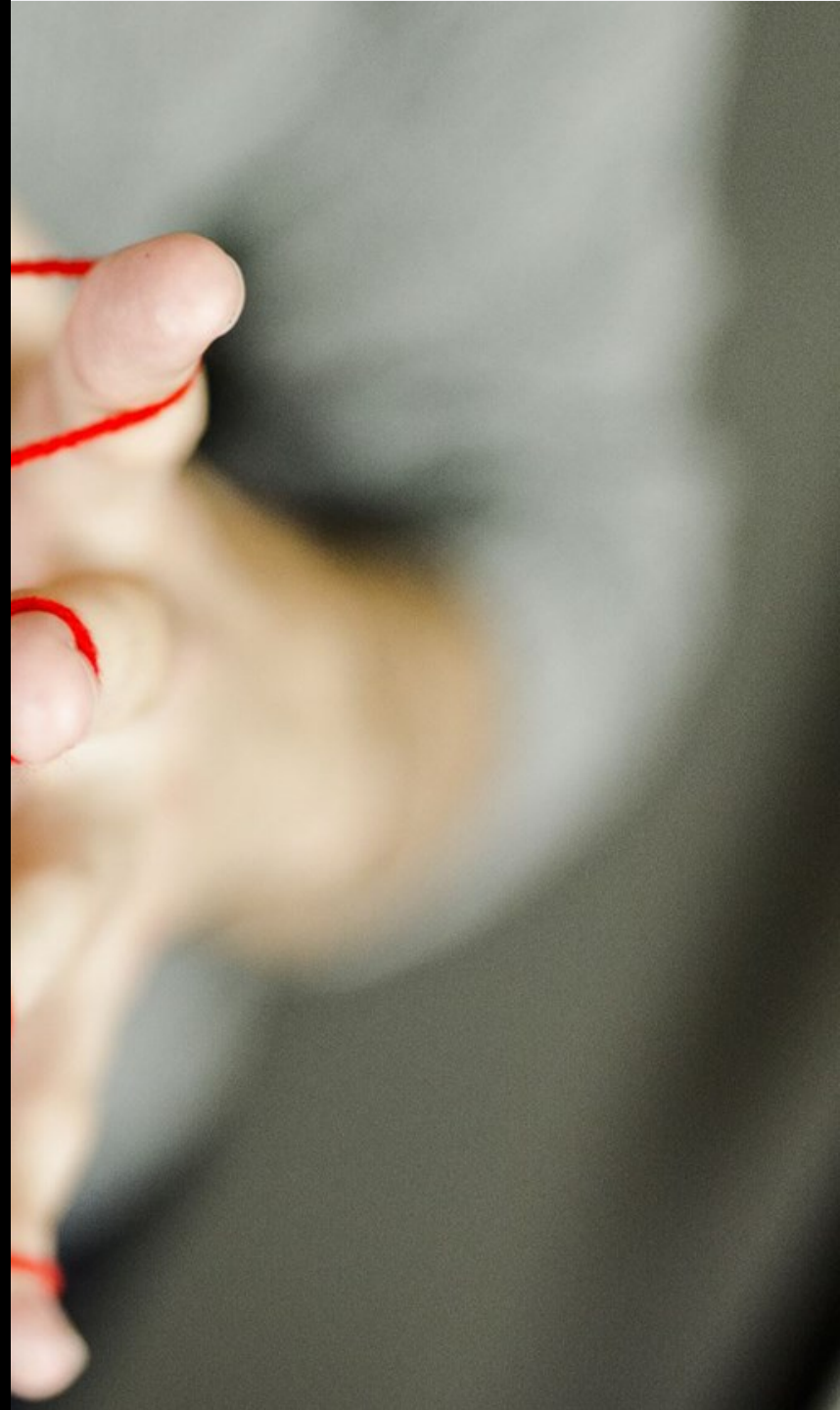
XGBOOST

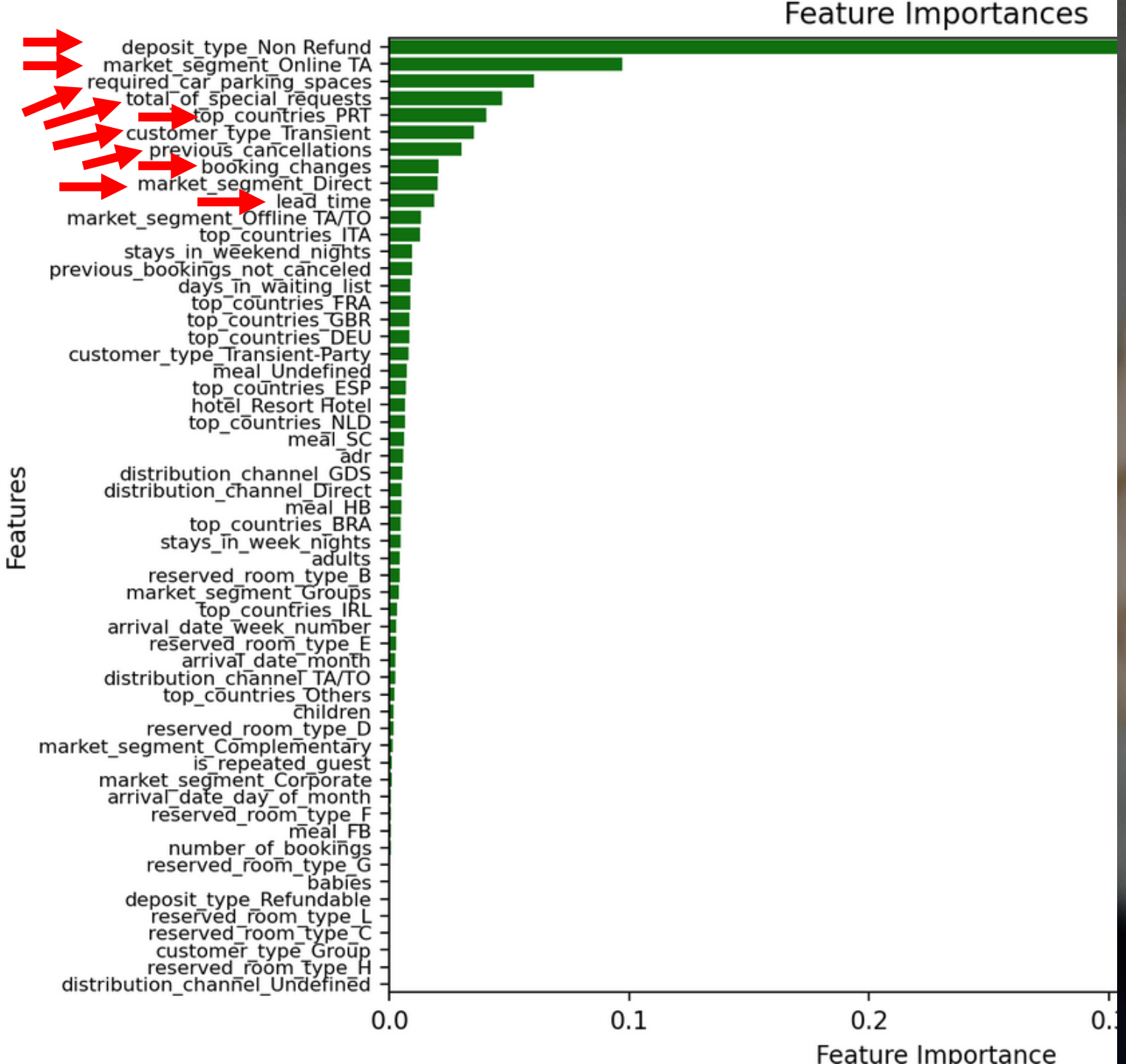
XGBOOST

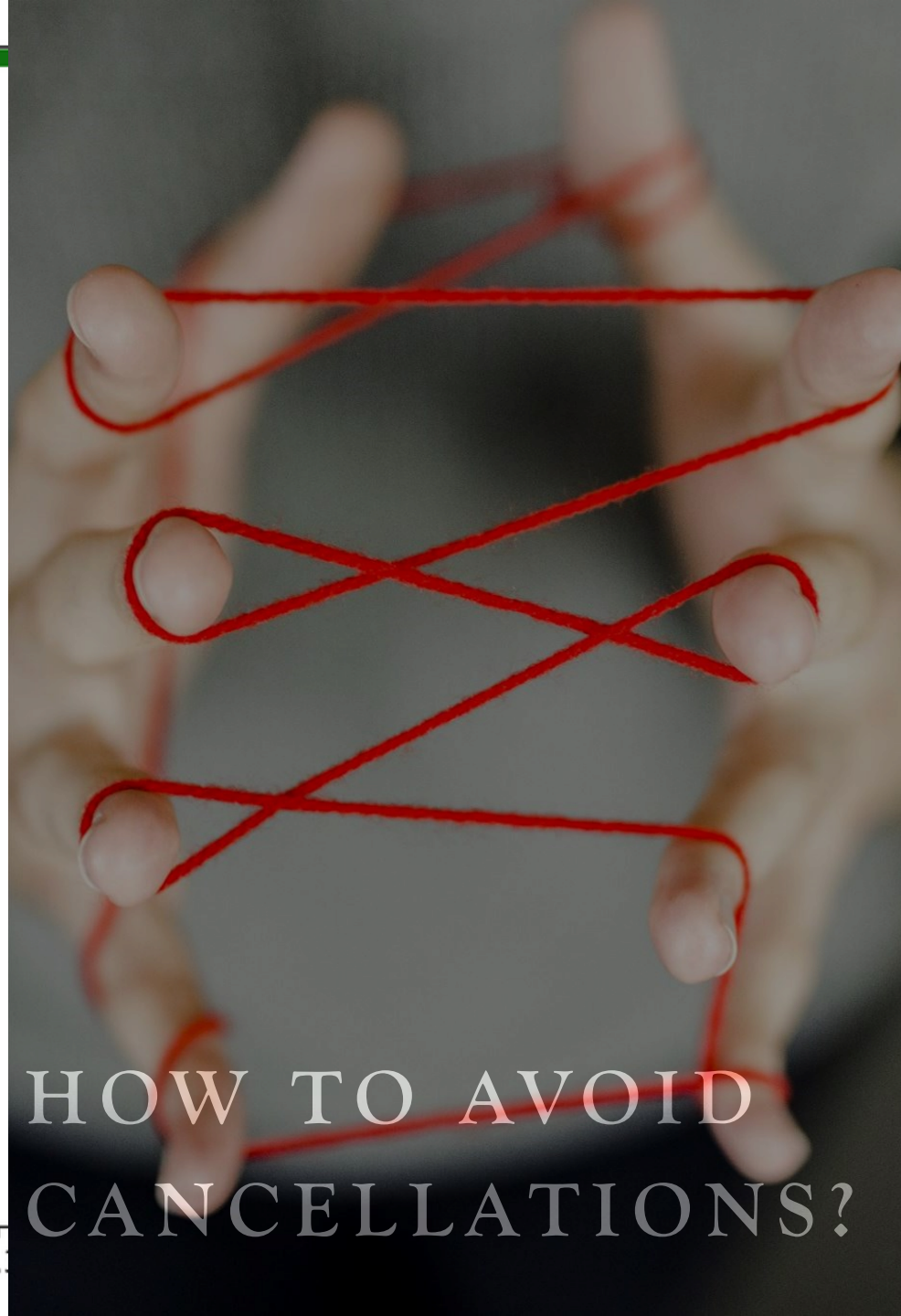
1. Accuracy
2. Cancellation Prediction Precision
3. Cancellation Recall
4. F1-Score for Cancellations

	precision	recall	f1-score	support
0	0.88	0.92	0.90	14651
1	0.85	0.79	0.82	8785
accuracy			0.87	23436
macro avg	0.87	0.85	0.86	23436
weighted avg	0.87	0.87	0.87	23436

HOW TO AVOID CANCELLATIONS?




Feature Importances
deposit_type_Non Refund
market_segment_Online TA
required_car_parking_spaces
total_of_special_requests
top_countries_PRT
customer_type_Transient
previous_cancellations
booking_changes
market_segment_Direct
lead_time
market_segment_Offline TA/TO
top_countries_ITA
stays_in_weekend_nights
previous_bookings_not_canceled
days_in_waiting_list
top_countries_FRA
top_countries_GBR
top_countries_DEU
customer_type_Transient-Party
meal_Undefined
top_countries_ESP
hotel_Resort Hotel
top_countries_NLD
meal_SC
adr
distribution_channel_GDS
distribution_channel_Direct
meal_HB
top_countries_BRA
stays_in_week_nights
adults
reserved_room_type_B
market_segment_Groups
top_countries_IRL
arrival_date_week_number
reserved_room_type_E
arrival_date_month
distribution_channel_TA/TO
top_countries_Others
children
reserved_room_type_D
market_segment_Complementary
is_repeated_guest
market_segment_Corporate
arrival_date_day_of_month
reserved_room_type_F
meal_FB
number_of_bookings
reserved_room_type_G
babies
deposit_type_Refundable
reserved_room_type_L
reserved_room_type_C
customer_type_Group
reserved_room_type_H
distribution_channel_Undefined
Features
0.0
0.1
0.2
Feature Importance


HOW TO AVOID
CANCELLATIONS?