

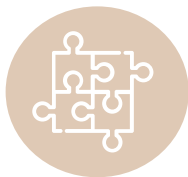
PREDICT HOTEL BOOKING CANCELATIONS

Q CONSULTING

TONGJIUZHOU LIU
AMELIE LANGENSTEIN
NINA URBANČIČ
LAURA CUNA



AGENDA



01
BUSINESS
UNDERSTANDING



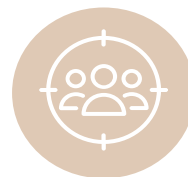
02
EXPLORATORY
DATA ANALYSIS



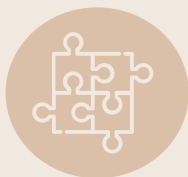
03
DATA
PREPARATION



04
MODELING



05
MODEL
DEPLOYMENT



01
BUSINESS
UNDERSTANDING



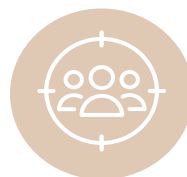
02
EXPLORATORY
DATA ANALYSIS



03
DATA
PREPARATION



04
MODELING



05
MODEL
DEPLOYMENT

01

BUSINESS UNDERSTANDING

THE PROBLEM



2018 cancellation rate 40%



free cancellation policies



dependency on weather
conditions



illness of guests



business meeting changes



deal seeking customers



online travel agencies OTAs



overbooking

01

BUSINESS
UNDERSTANDING

BUSINESS OBJECTIVE

Adjust offer
according to net
demand



Reduces risks of
overbooking or
vacant rooms



reduce cancellation to 20%



Minimize
costs



Maximize
revenue



Improve
reputation

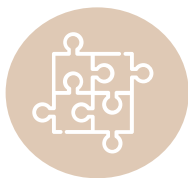
01

BUSINESS UNDERSTANDING

DATA MINING GOAL



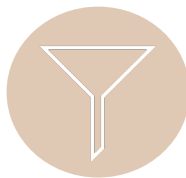
build a classification model that predicts if a booking will be canceled or not and predicts the probability of cancellation



01
BUSINESS
UNDERSTANDING



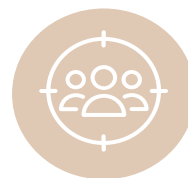
02
EXPLORATORY
DATA ANALYSIS



03
DATA
PREPARATION



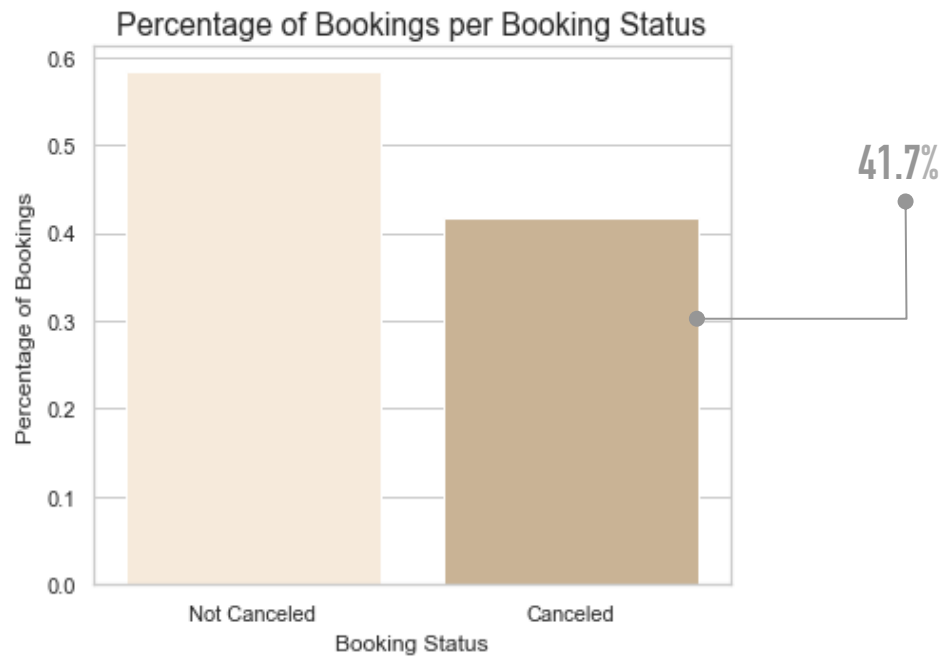
04
MODELING



05
MODEL
DEPLOYMENT

02

EXPLORATORY DATA ANALYSIS



02

EXPLORATORY
DATA ANALYSIS

WHERE ARE THE CUSTOMERS FROM?



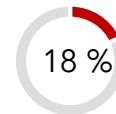
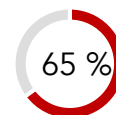
TOP 3 COUNTRIES
BOOKINGS

31K
PORTUGAL

8.8K
FRANCE

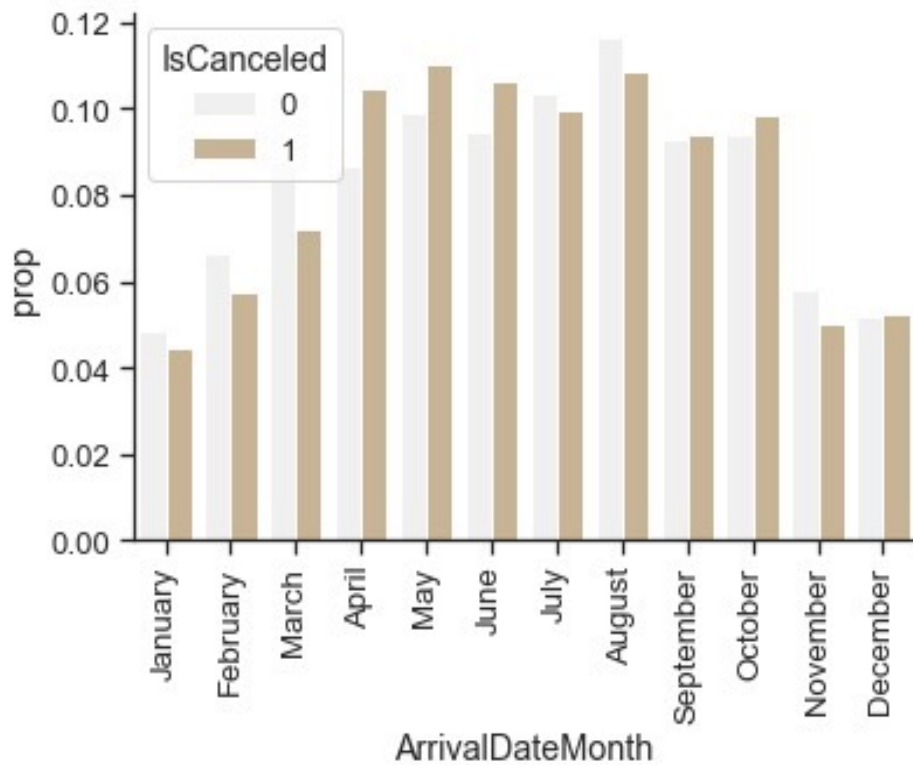
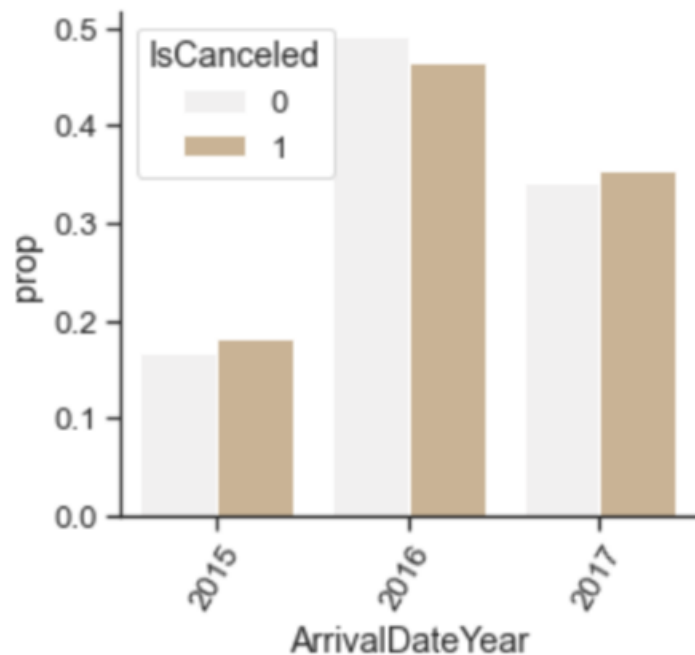
6.1K
GERMANY

CANCELATION RATES
OF THE TOP 3



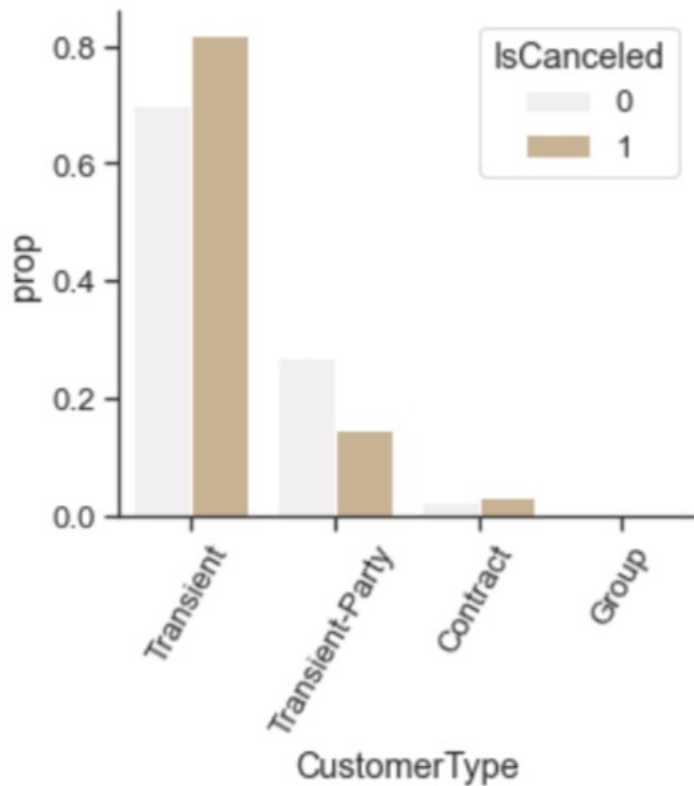
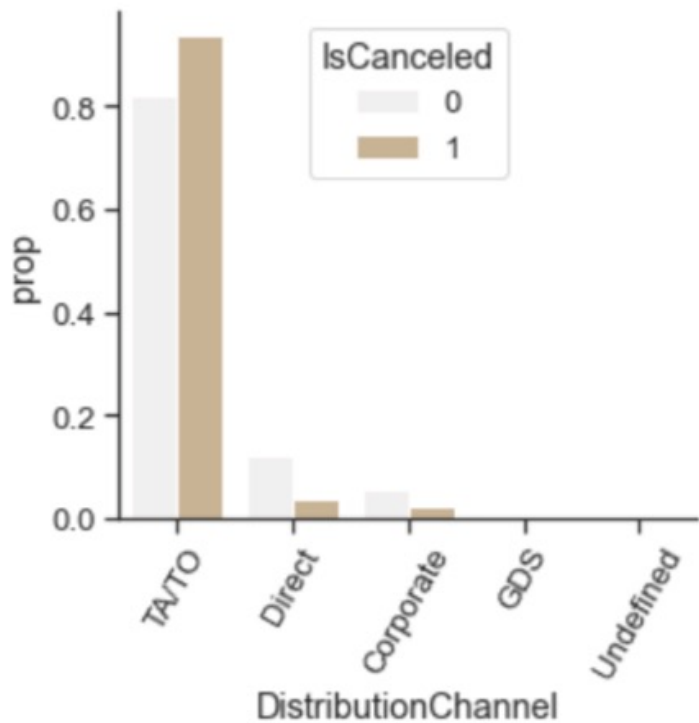
02

EXPLORATORY DATA ANALYSIS



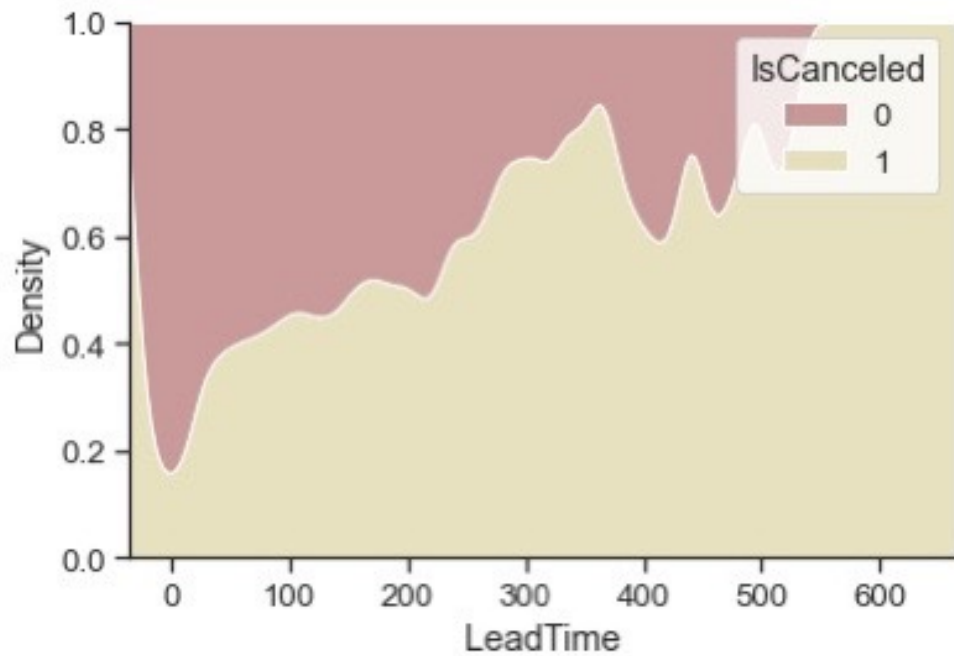
02

EXPLORATORY DATA ANALYSIS



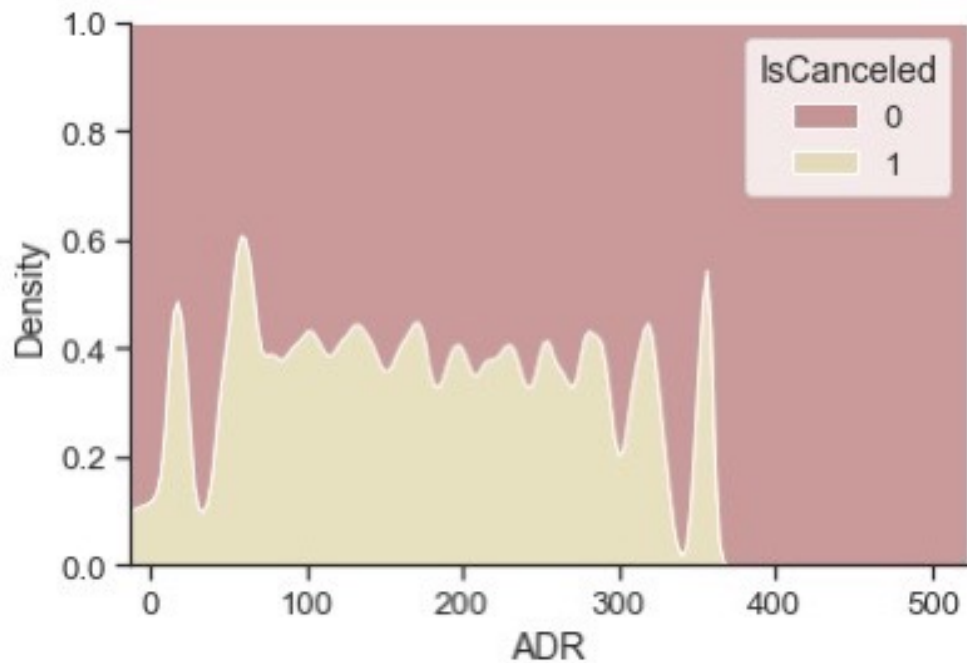
02

EXPLORATORY DATA ANALYSIS



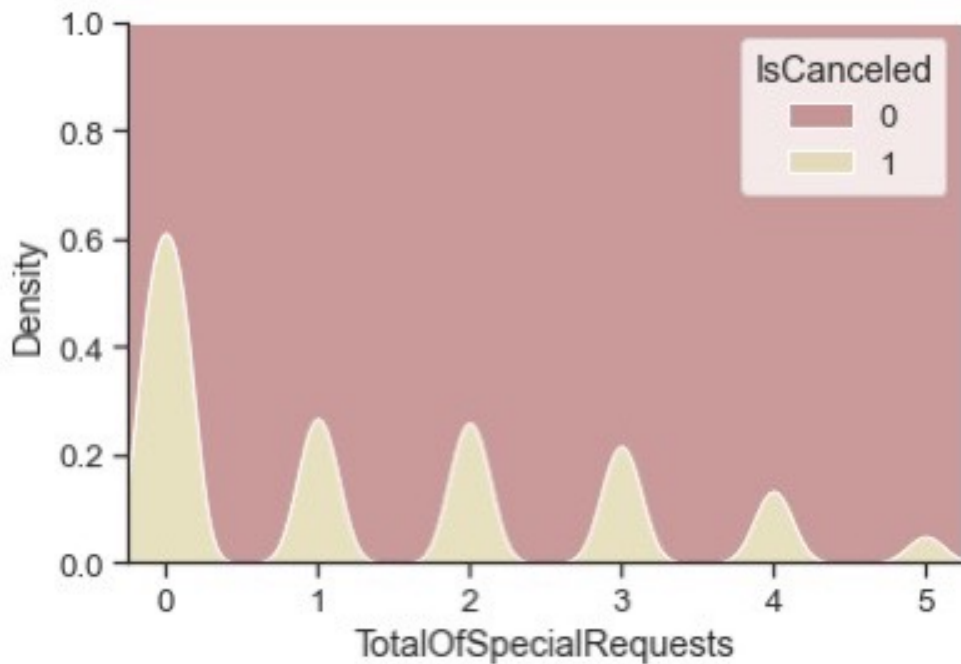
02

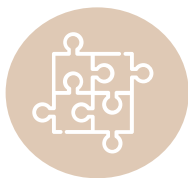
EXPLORATORY DATA ANALYSIS



02

EXPLORATORY DATA ANALYSIS





01
BUSINESS
UNDERSTANDING



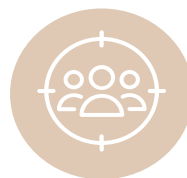
02
EXPLORATORY
DATA ANALYSIS



03
DATA
PREPARATION



04
MODELING



05
MODEL
DEPLOYMENT

03

DATA PREPARATION

Cleaned the data:

- dropped duplicates
- dropped irrelevant features
- substituted missing values with modes
- trimmed outliers
- encoded the categorical features with OneHotEncoding

Created new features:

TotalStayNights

StaysInWeekendNights + StaysInWeekNights

TotalGuestsNumber

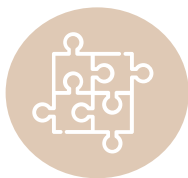
Adults, Children + Babies

ReservedRoomChanged

checking whether ReservedRoomType and AssignedRoomType were the same

TotalPreviousCancellationsRate

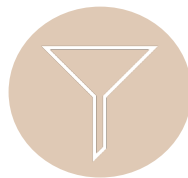
PreviousCancellations and total stays



01
BUSINESS
UNDERSTANDING



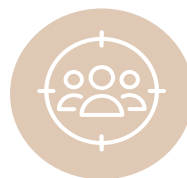
02
EXPLORATORY
DATA ANALYSIS



03
DATA
PREPARATION



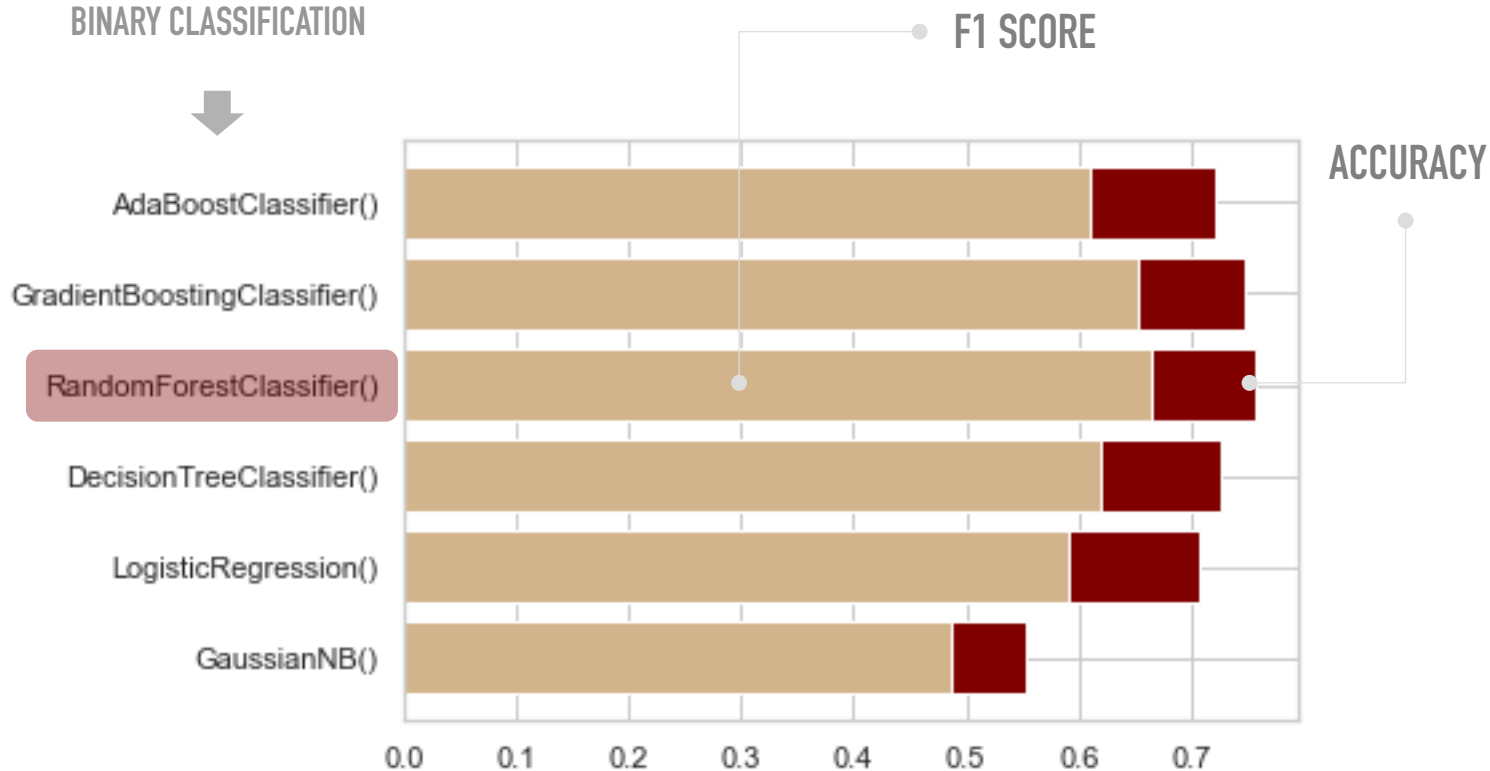
04
MODELING



05
MODEL
DEPLOYMENT

04 MODELING

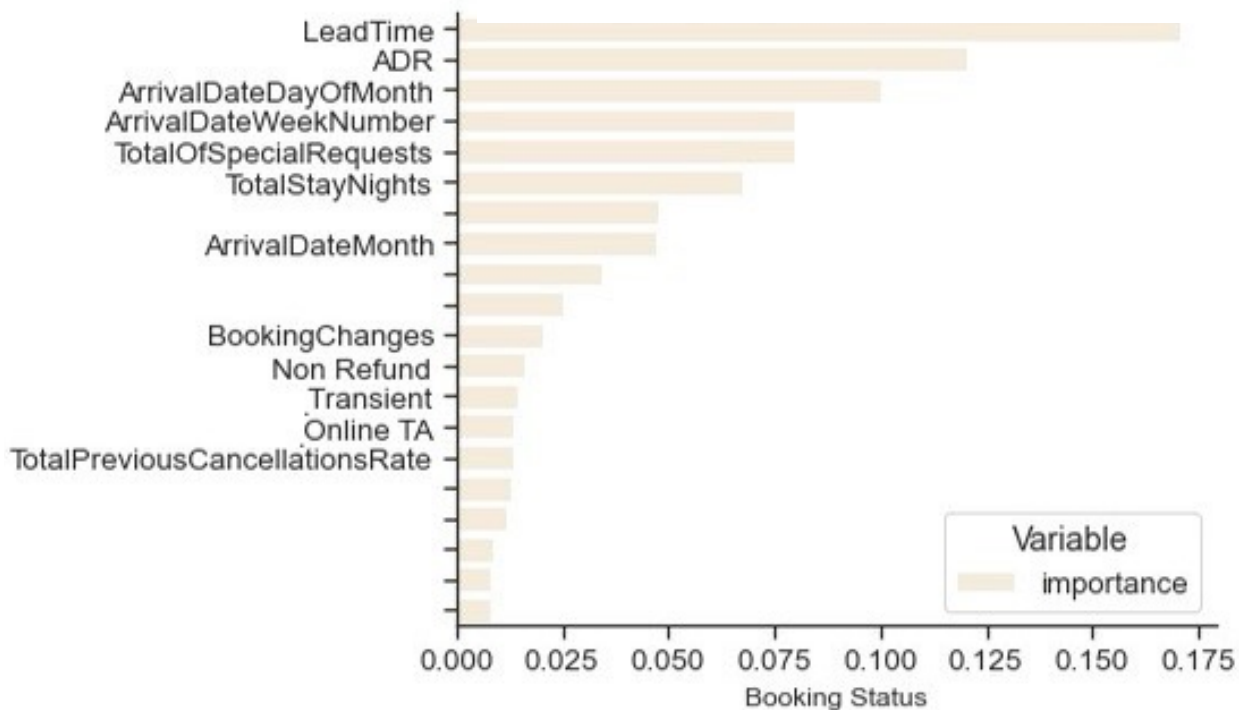
BINARY CLASSIFICATION



04

MODELING

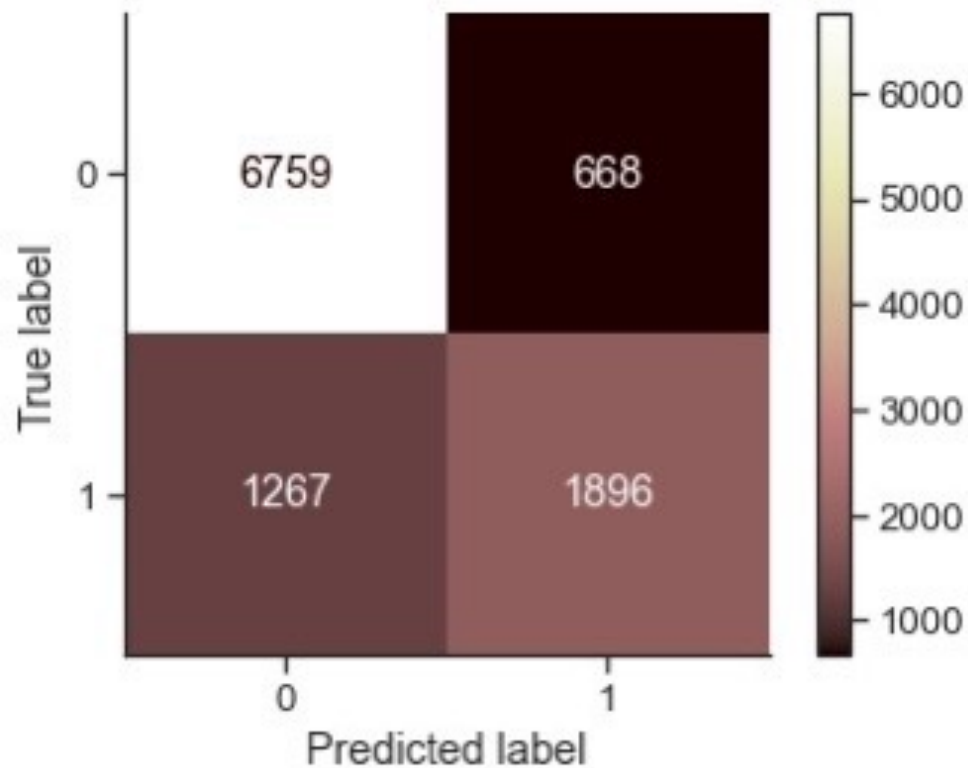
FEATURE IMPORTANCE

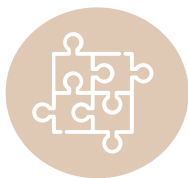


04

MODELING

CONFUSION MATRIX

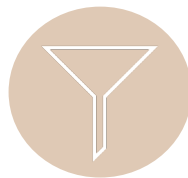




01
BUSINESS
UNDERSTANDING



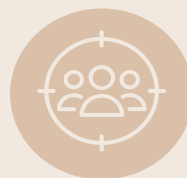
02
EXPLORATORY
DATA ANALYSIS



03
DATA
PREPARATION



04
MODELING



05
MODEL
DEPLOYMENT

05

MODEL DEPLOYMENT

MODEL

Random Forest
Classifier

- ✓ Predicts cancellation rate of future bookings

USE

Daily update of
cancellation rates

- ✓ estimate net demand
- ✓ Adjust overbooking measures
- ✓ Special offers

MAINTAIN

Train monthly with
new data

- ✓ Increase precision of model over time



For questions and concerns, contact us at
info@q-consulting.com

Copyright © 2022 Q CONSULTING. All rights reserved.