



Knowledge Miners

Hotel booking demand

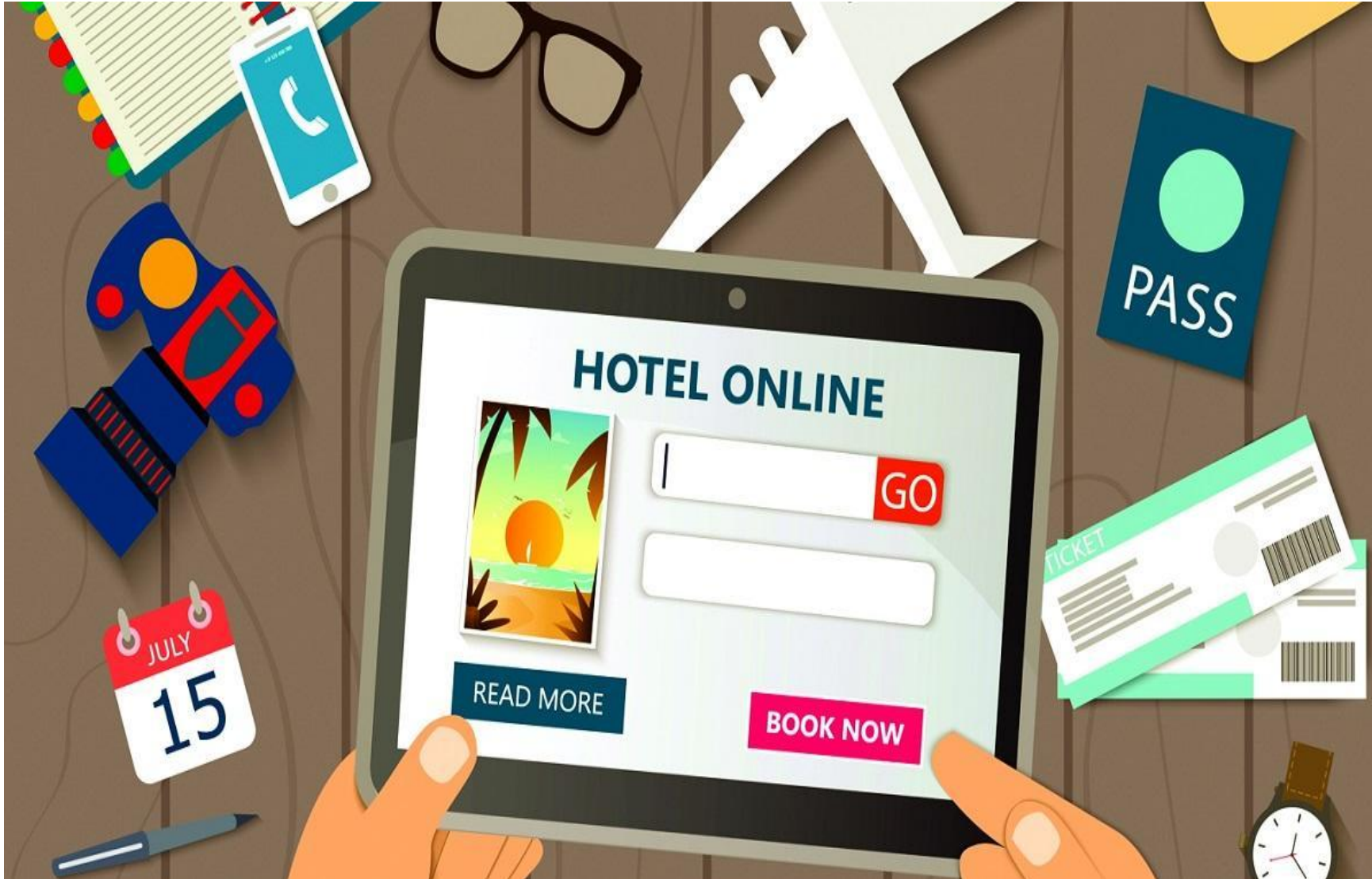
Suna Buse Agkoc, Siwoo Jung, Wei Tao Liu, Yannan Zhao

A modern hotel lobby with a curved bar, lounge chairs, and a large circular ceiling light fixture. The word "Hotel" is overlaid in large white text.

Hotel

AGENDA

- 01 | Introduction of Dataset
- 02 | Data Cleaning
- 03 | Model Building
- 04 | Model Comparison



Introduction of Dataset

32 columns

Integer

A String

Decimal

Other

17

13

1

1

Variable Name		Description
hotel		Hotel (H1 = Resort Hotel or H2 = CityHotel)
is_canceled		Value indicating if the booking was canceled (1) or not (0)
lead_time		Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
arrival_date_year		Year of arrival date
arrival_date_month		Month of arrival date
arrival_date_week_number		Week number of year for arrival date
arrival_date_day_of_month		Day of arrival date
stays_in_weekend_nights		Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
stays_in_week_nights		Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
adults		Number of adults
children		Number of children
babies		Number of babies
meal		Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal
country		Country of origin. Categories are represented in the ISO 3155–3:2013 format
market_segment		Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
distribution_channel		Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
is_repeated_guest		Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_cancellations		Number of previous bookings that were cancelled by the customer prior to the current booking
previous_bookings_not_cancelled		Number of previous bookings not cancelled by the customer prior to the current booking
reserved_room_type		Code of room type reserved. Code is presented instead of designation for anonymity reasons.
assigned_room_type		Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due
booking_changes		Number of changes/amendments made to the booking from the moment the booking was entered on the PMS
deposit_type		Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was
agent		ID of the travel agency that made the booking
company		ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for
days_in_waiting_list		Number of days the booking was in the waiting list before it was confirmed to the customer
customer_type		Type of booking, assuming one of four categories: Contract- when the booking has an allotment or other type of
adr		Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
required_car_parking_spaces		Number of car parking spaces required by the customer
total_of_special_requests		Number of special requests made by the customer (e.g. twin bed or high floor)
reservation_status		Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out
reservation_status_date		Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to

Target Audience

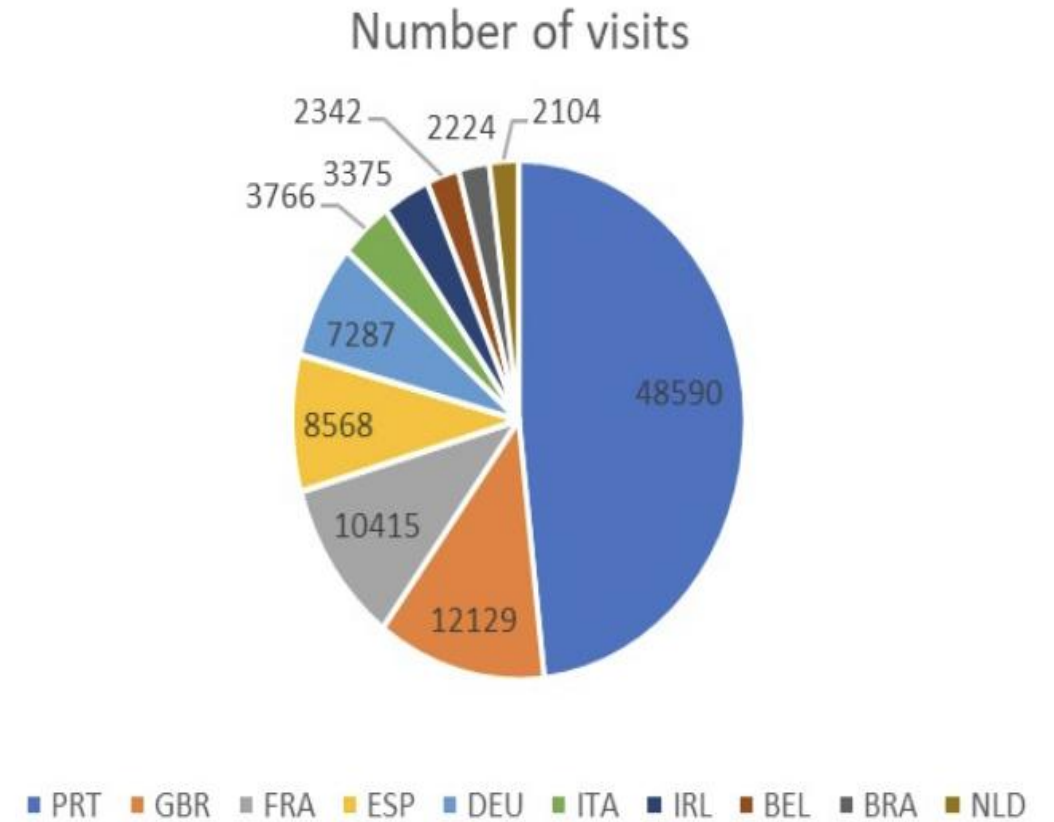
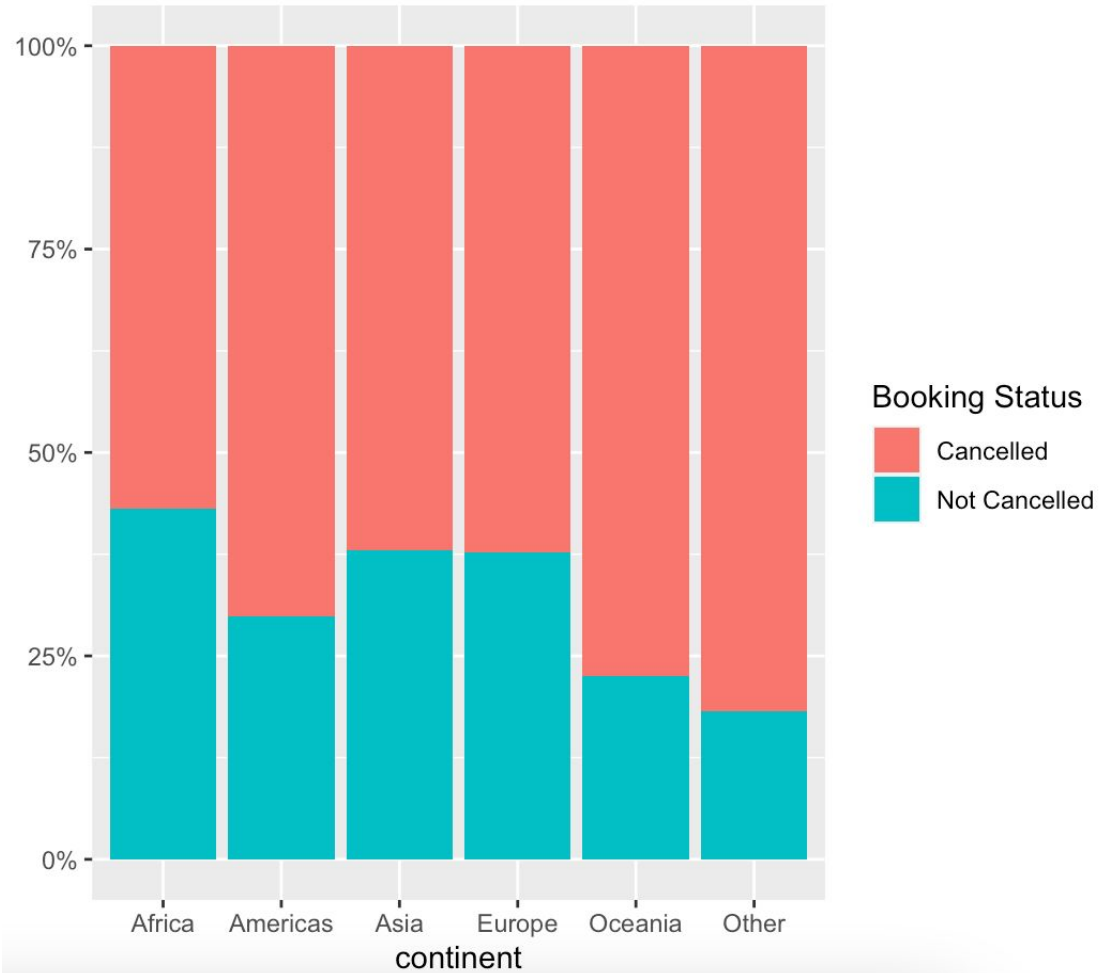
- **Marketer**
- **Hotel Owners**
- **Hotel Customers**
- **Travel Agencies**

Goal

Our goal is to find out a dummy variable that indicates whether the booking was canceled or not. We want to know which factors most impact the hotel booking demand/cancellation.

Part 2 Data Cleaning & Exploration

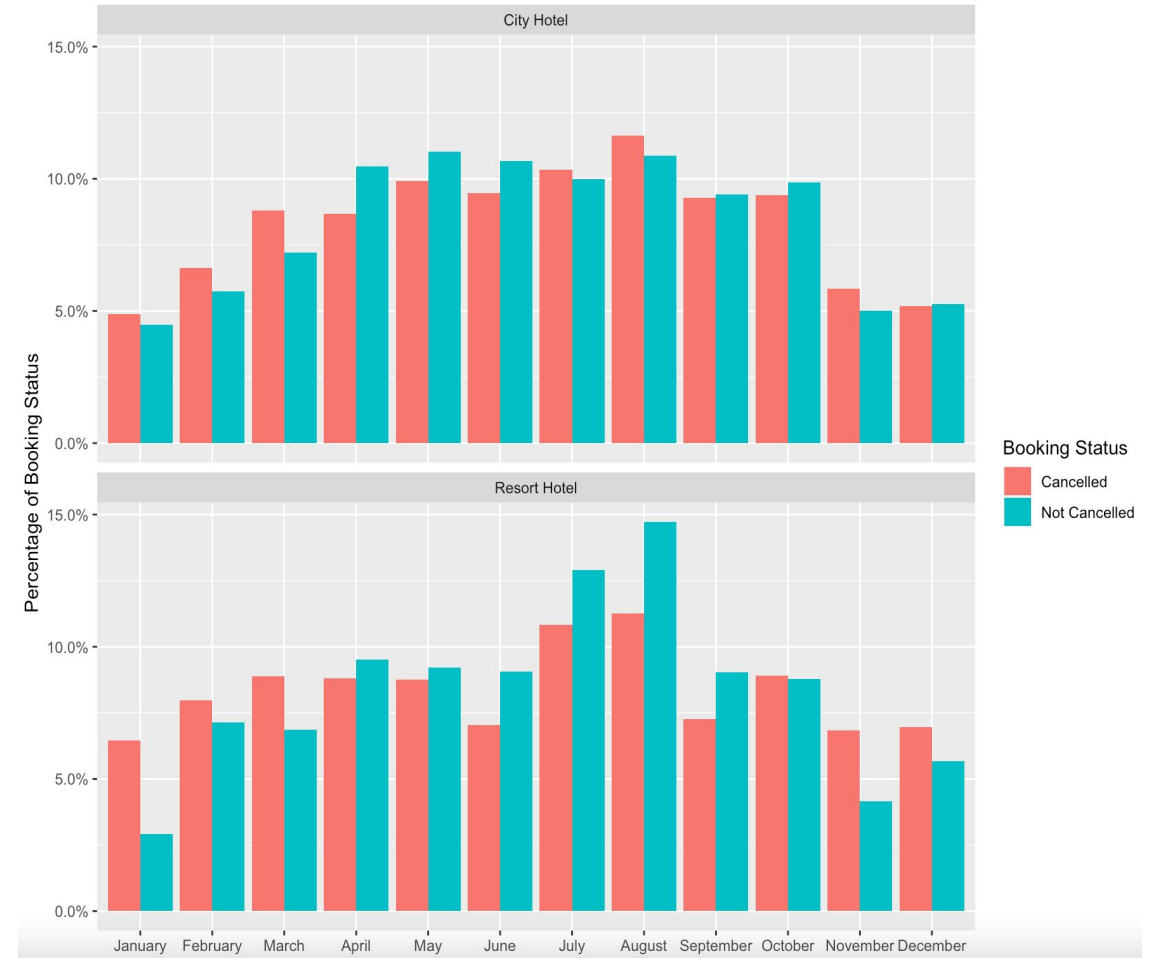
Hotel Customer



Top three countries of the guests are PRT, GBR and FRA.

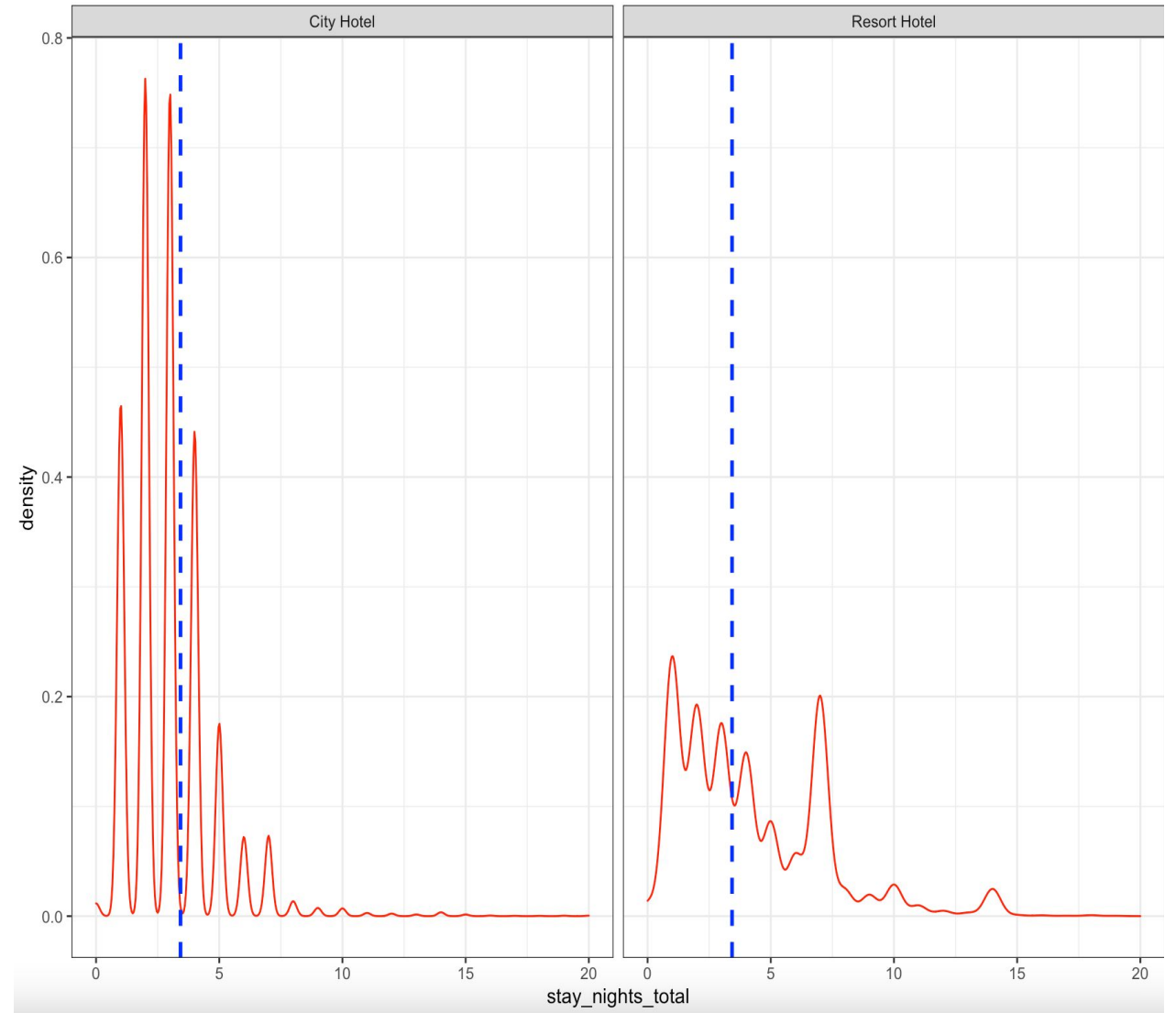
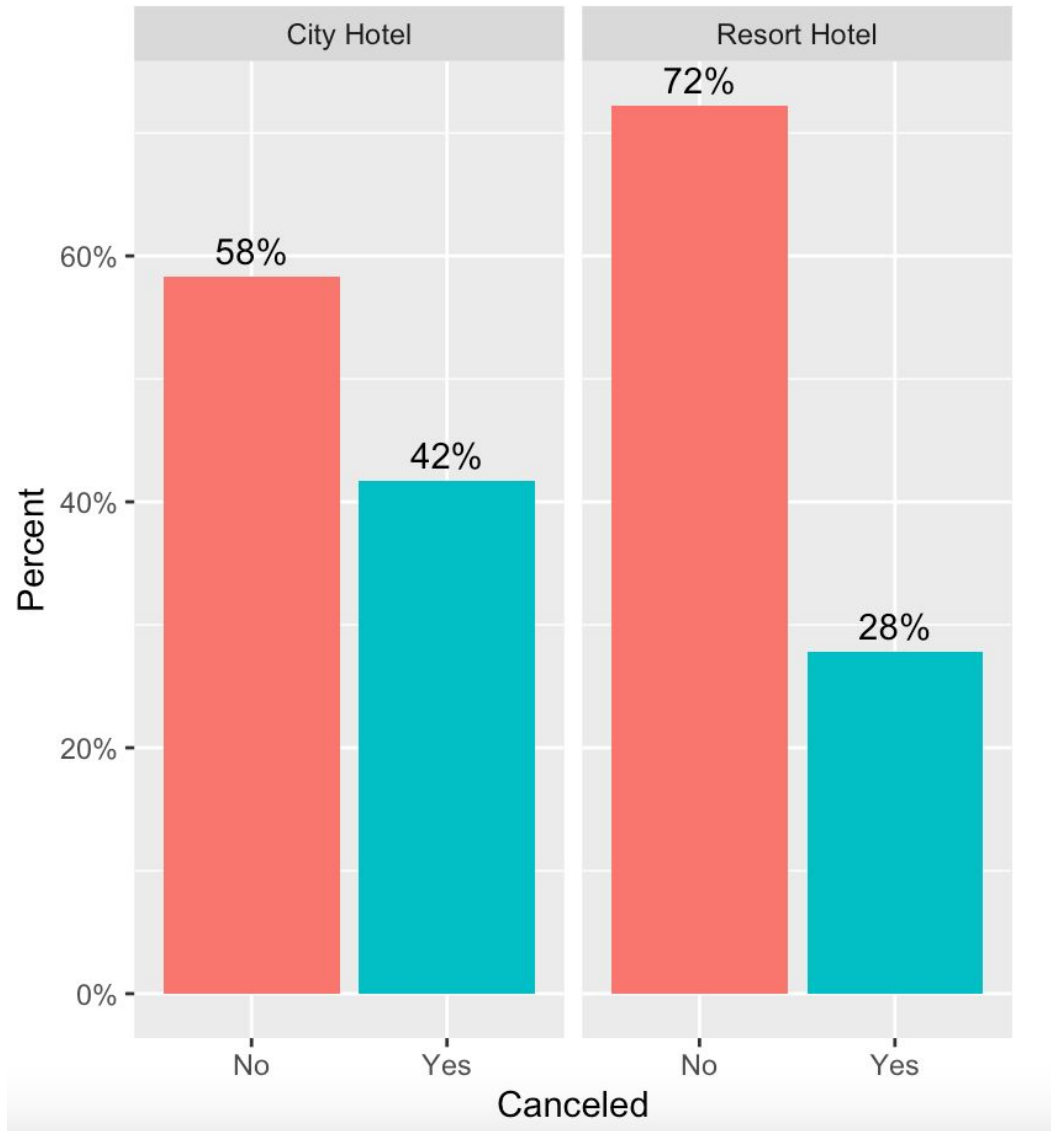
Part 2 Data Cleaning

Booked Year



May, July and August are the busiest three months

Hotel



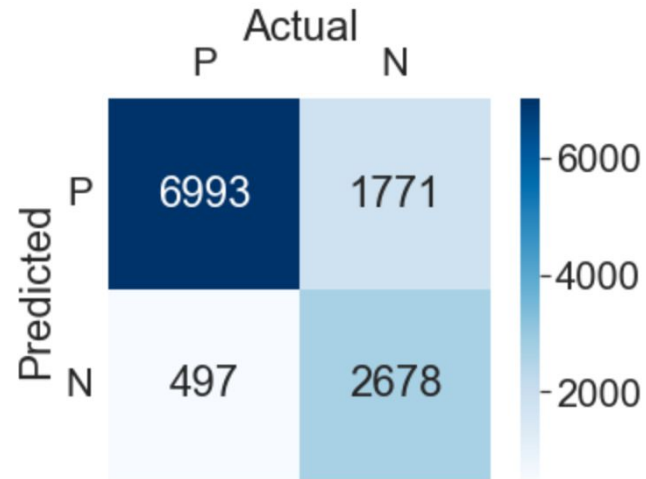
Classification Models

- **Logistic Regression**
- **Decision Tree**
- **Random Forest**

Key Points

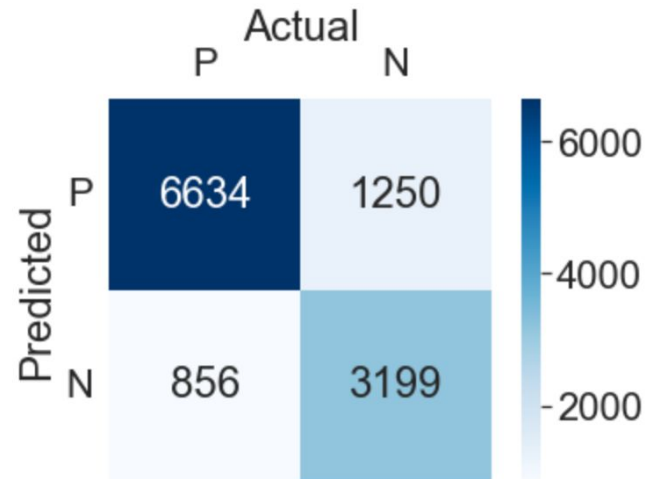
- **Split the Train/Test data by 80:20 ratio**
- **Shuffle the dataset**
- **Randomization**
- **Cross Validation (10th Fold)**
- **Hyperparameter Tuning**

Logistic Model



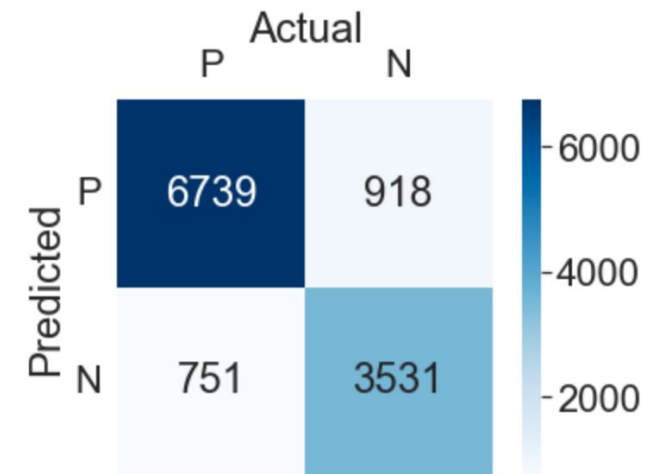
Accuracy : 0.81
Precision : 0.80
Recall : 0.93
F1 : 0.86

Decision Tree



Accuracy : 0.82
Precision : 0.84
Recall : 0.89
F1 : 0.86

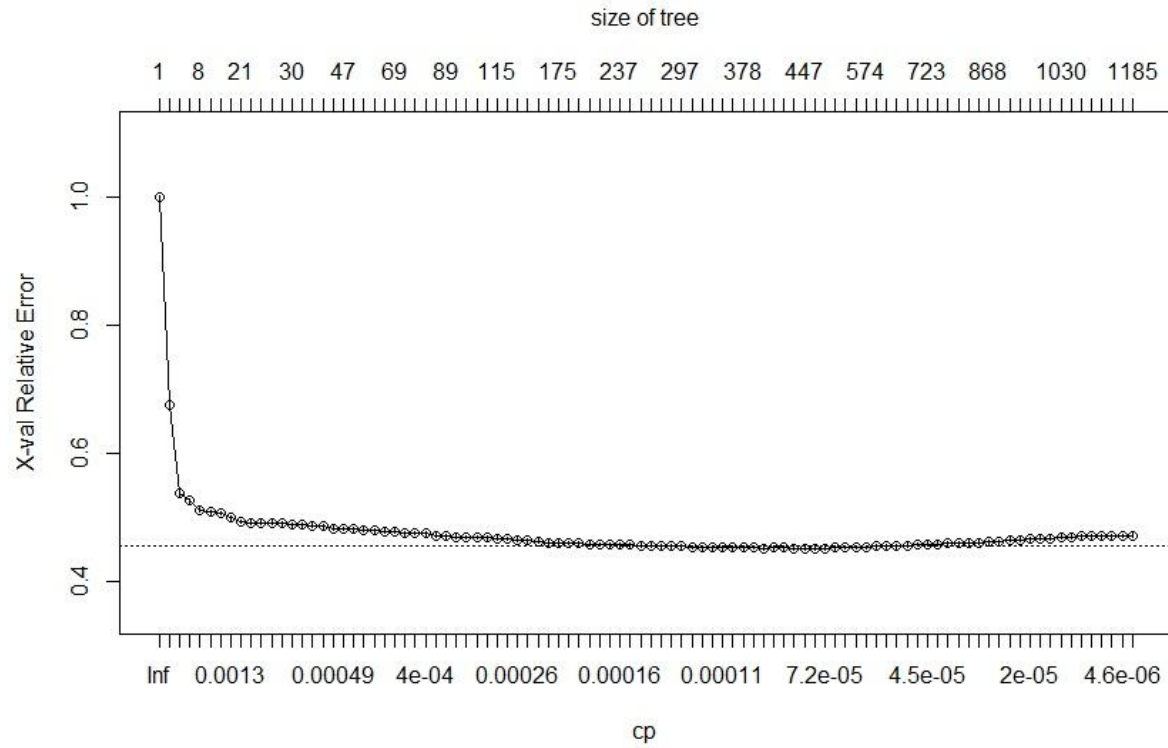
Random Forest



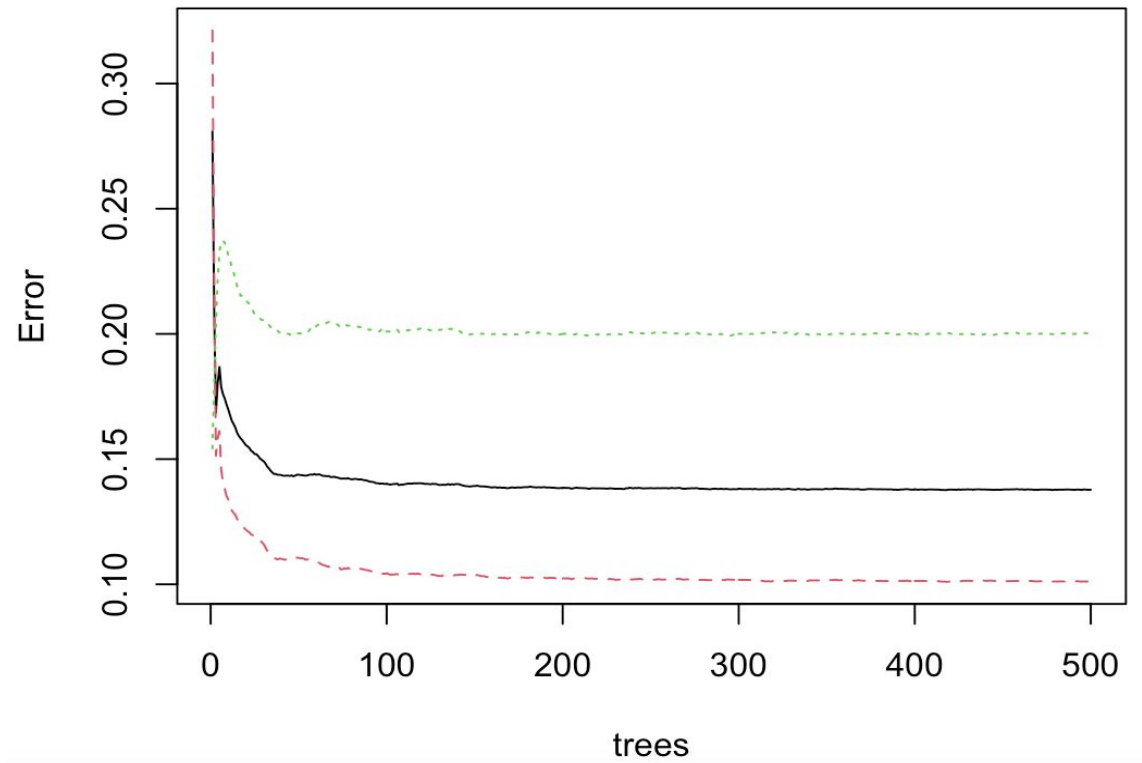
Accuracy : 0.86
Precision : 0.88
Recall : 0.89
F1 : 0.88

Part 3 Error Term and Tuning

Decision Tree X-val Relative Error



Error rate of random forest



Part 3 Tuning and ROC

Hyperparameters:

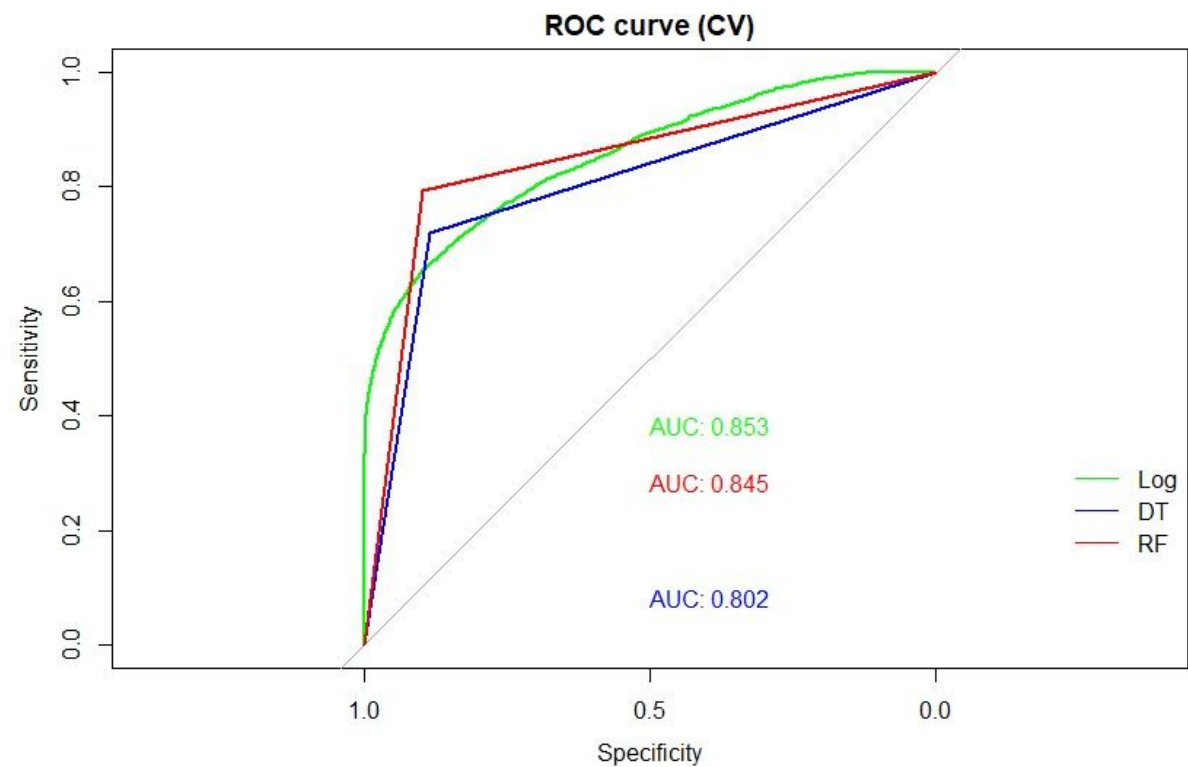
Maxdepth: Depth the tree will grow to, the deeper, the more complex, but less training errors.

Cutoff: Threshold to define a probability of a positive event

Mtry: Number of variables or features to randomly sample for each split

Sampsize: How many units to sample from each class

Model	Accuracy	Precision	Recall	F1 Score
Log	0.8100	0.7979	0.9336	0.8605
DT	0.8236	0.8415	0.8857	0.8630
RF	0.8589	0.8792	0.8985	0.8887

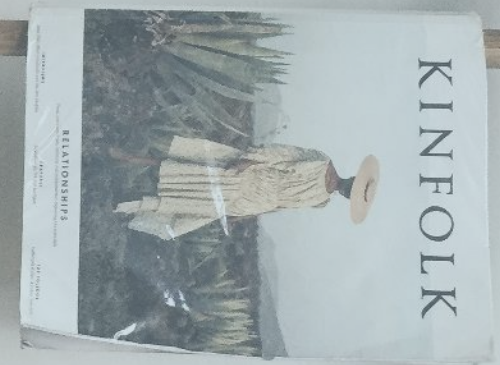


Reference

s

[Hotel booking demand | Kaggle](#)

[Hotel booking demand datasets – ScienceDirect](#)



The background of the image shows a modern building with a glass facade. The building is partially obscured by horizontal blinds, which are closed, creating a pattern of light and shadow. The blinds are white and run diagonally across the frame. The building's glass reflects the sky and some greenery. The overall tone is muted and professional.

Thank you