DUKE FUQUA SCHOOL OF BUSINESS

Hotel Booking Demand Prediction

Section C Team #61

Anoohya Gaddam, Collin LaButte, Maggie Guo, Steven He, Wei-Ting (Katherine) Chang

10.23.23

Business Understanding

Business Problem	Data Mining Solution	Business Value
• Cancellations lead to revenue loss, poor utilization, and operational efficiency problems • In the dataset, 37%were canceled before the arrival date accounting for 16 M euros of lost revenue	Predict whether a customer will cancel the hotel booking or not	Increase efficiency and profit by overbooking

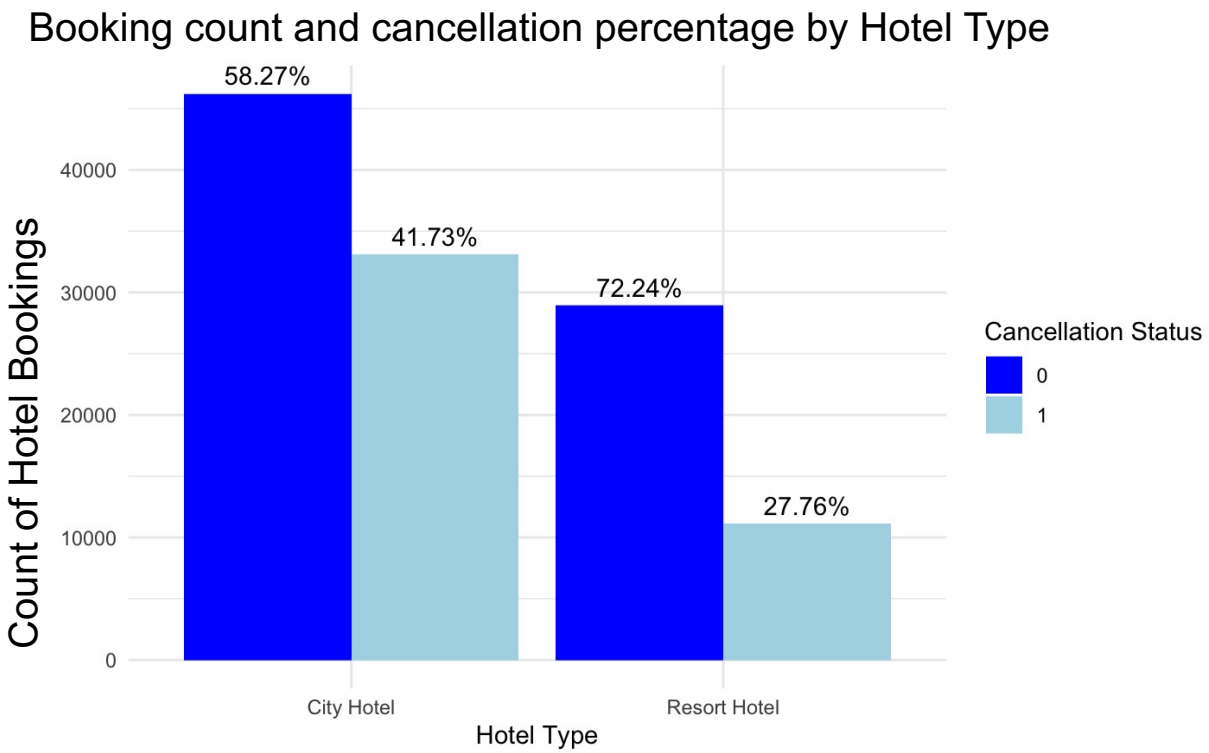
Data Understanding

- The dataset contains hotel bookings from various countries between 2015 – 2017

- Features/Attributes
  - Demographics: country, customer type, adults, children, babies,...
  - Booking Information: lead time, arrival date week number, previous cancellations, previous bookings not canceled, booking changes, ADR(Average Daily Rate),...
  - Services: meal, market segment, distribution channel, ...
  - Hotel type: hotel
- Target
  - is_canceled (0/1)

DUKE FUQUA SCHOOL OF BUSINESS

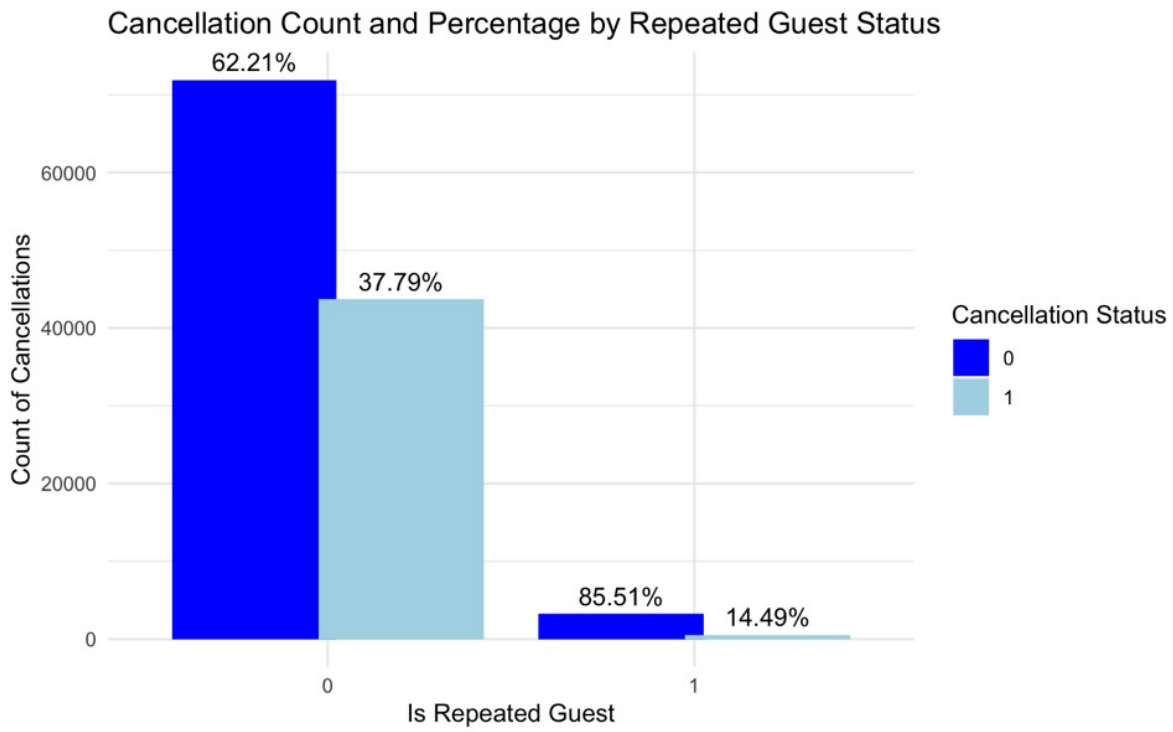
Data Visualization



- More bookings in a city hotel when compared to the resort, cancellation percentage is also higher.
- Resort hotels: customers usually planned in advanced for traveling purpose and has lower possibilities to cancel

DUKE FUQUA SCHOOL OF BUSINESS

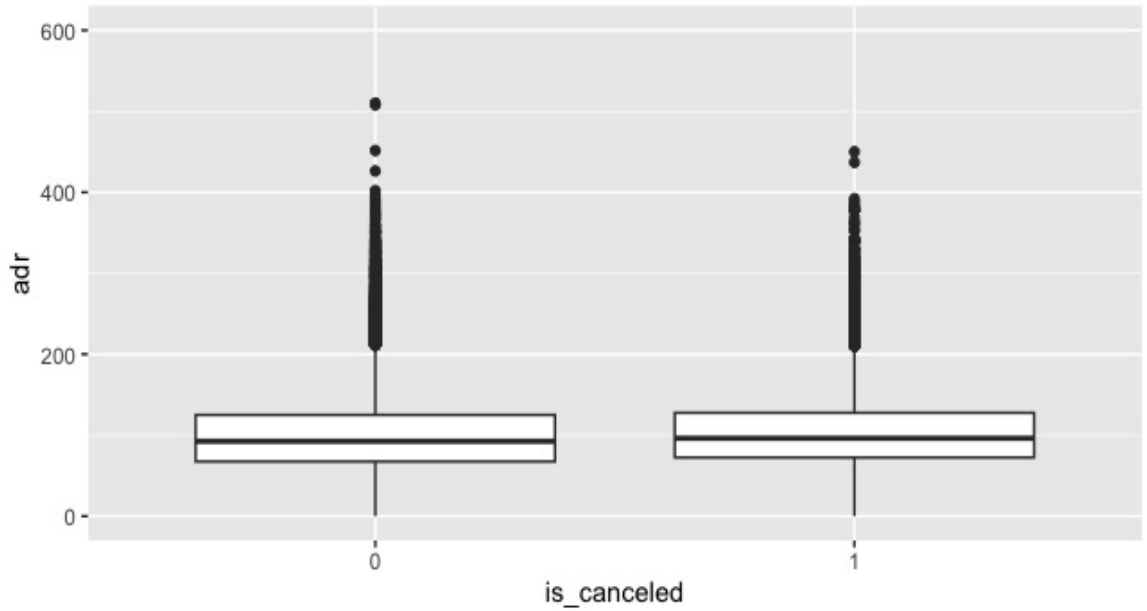
Data Visualization



- Repeated guests are less likely to cancel the bookings. This might be because the customers are loyal, or they might be frequent visitors for work purposes

DUKE FUQUA SCHOOL OF BUSINESS

Data Visualization



- Hotel's average daily rate didn't significantly affect the cancellation action.

DUKE FUQUA SCHOOL OF BUSINESS

Modeling

- Binary prediction: Utilize supervised learning models for classification purpose
- Cross-validation: K-fold cross-validation
- The performance metrics: OOS Accuracy

Model	OOS Accuracy
Logistic Regression	**81.43%**
SVM	**82.84%**
Classification Tree	**80.30%**
XGBoost	**84.09%**

- Prediction Result: 84.378%

DUKE FUQUA SCHOOL OF BUSINESS

K-means Clustering

	is_ canceled	lead_time	arrival_date_day _of_month	adults	children	babies	is_repeated_ guest	adr
1	0.348	43.285	16.159	2.020	0.278	0.012	0.011	166.994
2	0.249	31.320	15.562	1.718	0.043	0.007	0.0642	75.273
3	0.451	160.728	16.040	1.911	0.096	0.007	0.004	103.189
4	0.644	326.526	15.488	1.945	0.047	0.003	0.015	85.041

DUKE FUQUA SCHOOL OF BUSINESS

Deployment

- The results from the data mining should be used to predict whether the customer cancel the booking or not
- By calculating the percentage of booking rate minus the cancellation, the hotels would be able to release certain percentage of rooms for overbooking
- Mitigate the risk of overbooking problems by setting conservative overbooking thresholds and partnering with other local hotels

DUKE FUQUA SCHOOL OF BUSINESS