Applying machine learning to hotel booking cancellation prediction

Business Analytics Final Project

Shiling Chen, Wenyu Han, Yumeng Tian,
Jiaming Xie, Yuxuan Zhang, Ziyang Zhang

Objective & Background

How can hotels save $1 M?

- Why upgrade
- Overselling
- Overselling rate
- Cancellation rate

How to estimate the cancellation probability for each customer?

Data Overview

Date Information:

reservation date, lead time, check-out date...

Booking Information:

Room type, Meal, Booking Channel, Market Segment, Deposit Type...

Personal Information:

of adults, children

Data Visualization

Correlation

Positive Correlated Features:

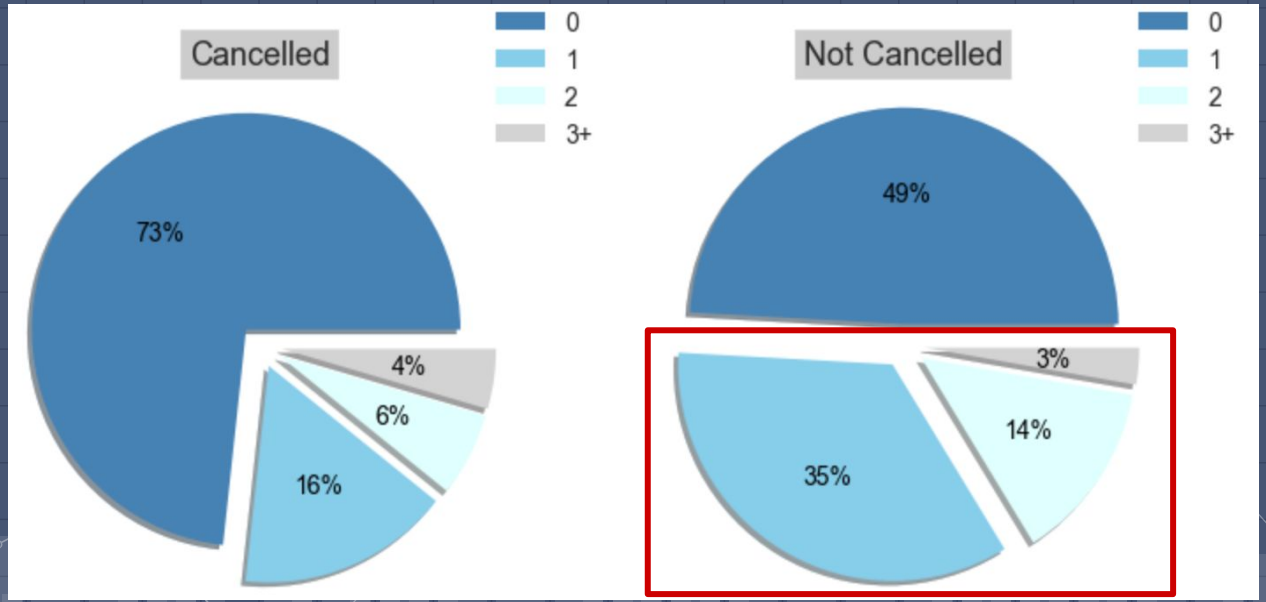
- lead time
- previous cancel rate
- previous cancellations

Negative Correlated Features:

- total of special request
- required car parking spaces
- booking changes

Data Visualization

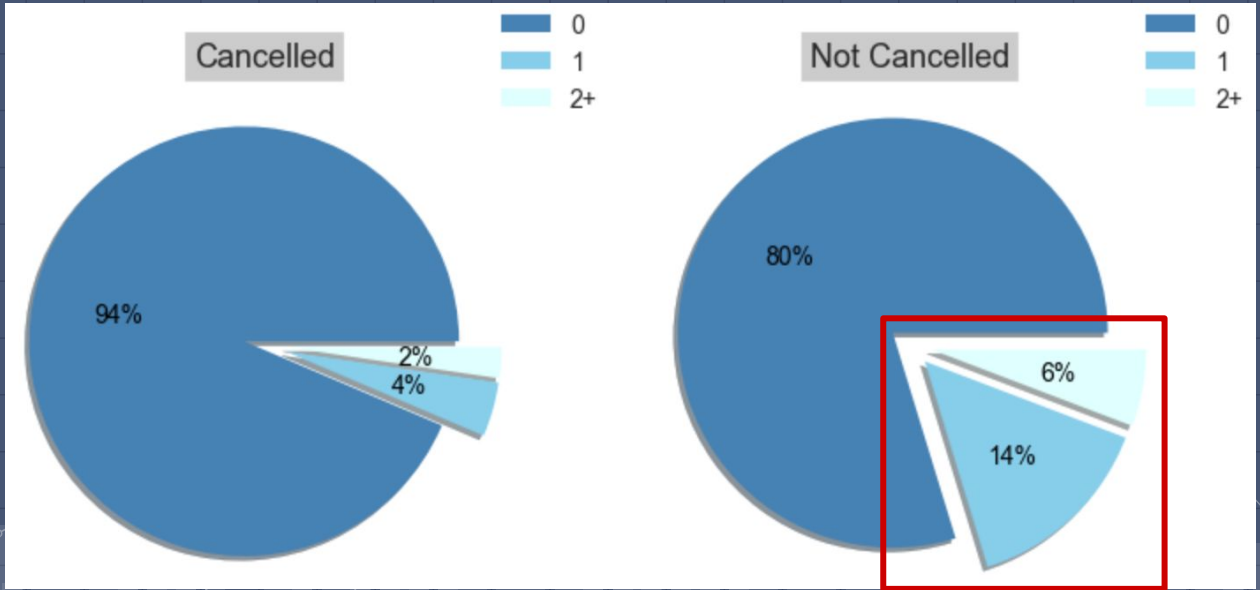
Special Requests



- For not cancelled orders, guests have significantly more special requests.
- Bookings with requests may be less likely to be cancelled.

Data Visualization

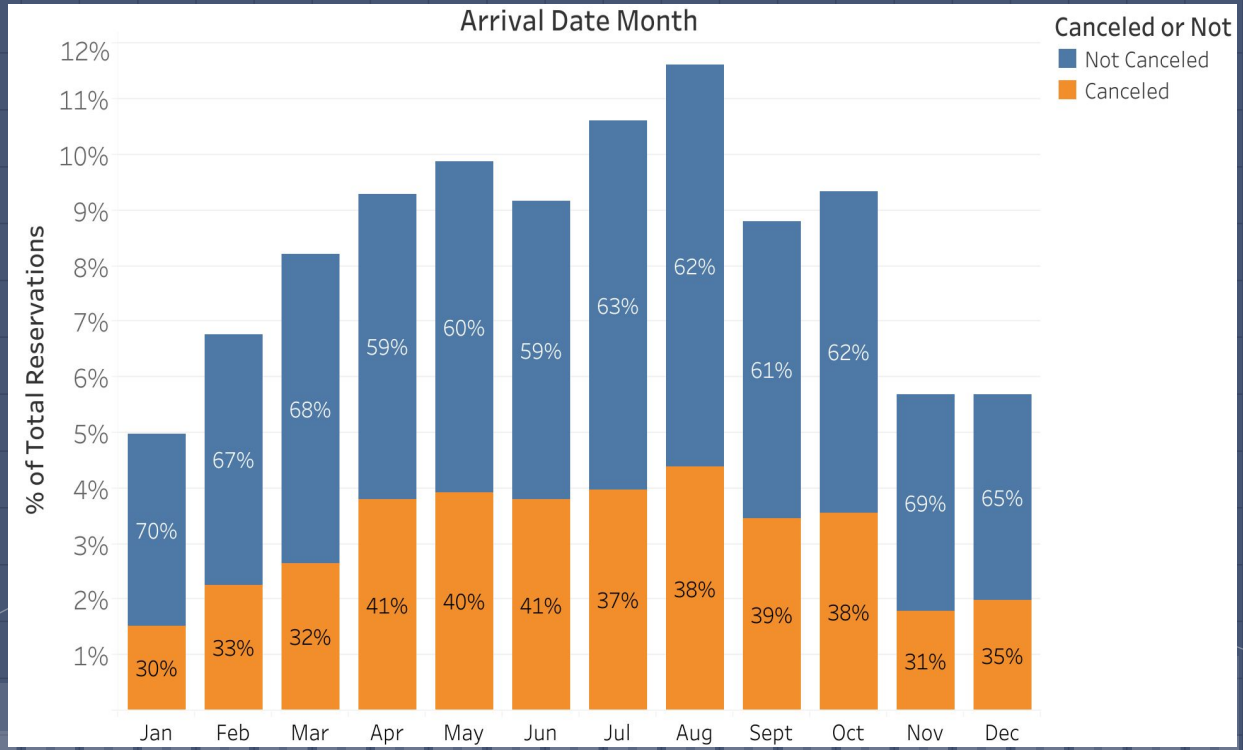
Booking changes



- Changed bookings take up a higher percentage in not cancelled orders than in cancelled orders.
- Orders with more than one booking changes tends to have lower cancellation probability.

Data Visualization

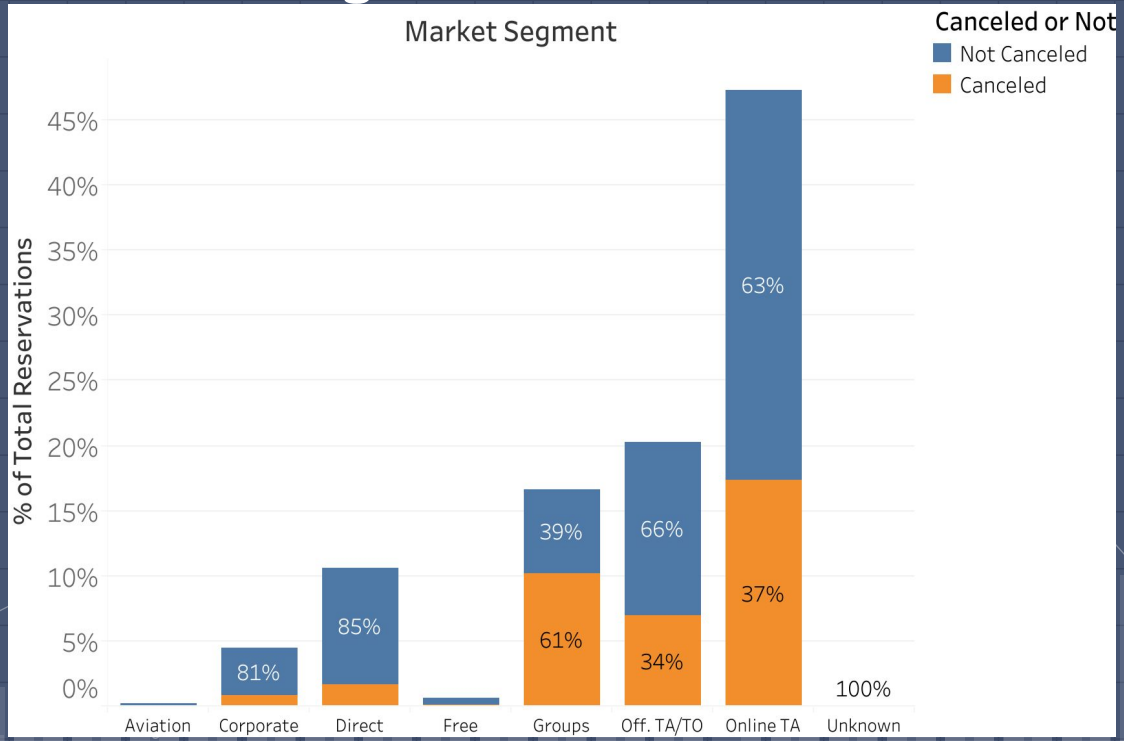
Seasonality



- The seasonality of the hotel bookings does exist.
- Winter months(Nov. To Mar.) have much fewer reservations.
- The cancellation rate of each month decreases as total reservations goes down.

Data Visualization

Market Segment



- Travel Agents/Tour Operators segment weighs over 50% of the total bookings.
- Direct and Corporate segments have lower cancelation rates compared to TA/TO.
- Groups segment has an extremely high cancelation rate over 60%.

Cost Model Construction

				
Average rate per room		Average gross profit per room		Average cost per empty room
$141	−	$71	=	$70

- Data from HotStats' report.
- Major Assumptions:
  - No additional consumptions like foods, facilities, etc.
  - The opportunity cost of an empty room ≈ The operation cost of an occupied room
  - Total gross profit of a hotel = # of empty room *(-70) + # of occupied room *71

Cost Model Construction

Before

After

Correct Prediction
-> no empty room
-> +$71

Those who do not cancel are predicted to cancel
-> need to find another room
+0.2*(71*2) -> find an empty room under the same hotel chain
+0.8*(71-141-141*0.1) ->find another room of another hotel
-0.05*141 ->reputation loss
= -$46

Predict

	Actual	Not Cancel	Cancel
$71	Not Cancel	(TN) Room Occupied Properly	(FP) Need to find a new room...
-$70	Cancel	(FN) Empty Room...	(TP) Room Occupied Properly

Those who cancel are not recognized by the model
-> empty room
-> -$70

Correct Prediction
-> no empty room
-> +$71

Machine Learning Results

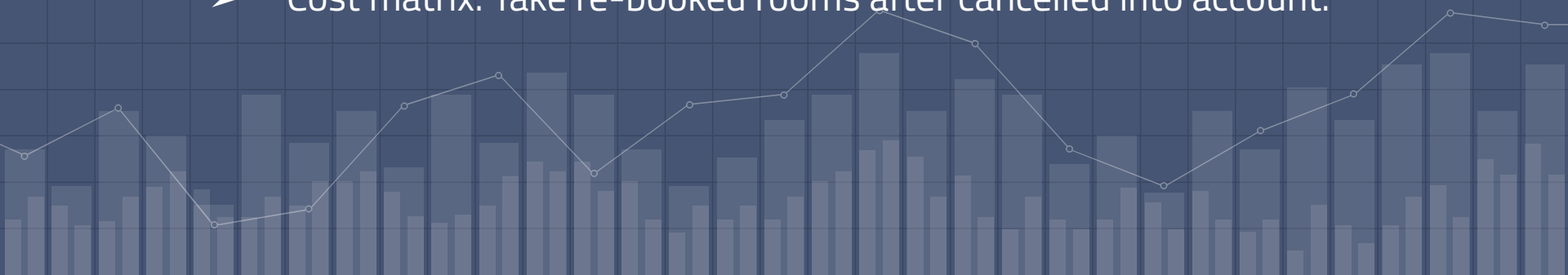
	Accuracy	Precision	Recall	Additional Profit Gain per Room
Logistic regression	79%	72%	72%	$26
Decision tree	79%	78%	62%	$25
XGboost	84%	85%	70%	$32
Random forest	85%	85%	73%	$33
Neural network	86%	83%	79%	$34

Total profit gain:
$1,017,828

*Test size = 0.25; 29847

Advice for Hotels

- Advice For Hotels:
  - Overbooking rooms is better than leaving it empty;
  - Use *Neural Network* to predict the cancellation rate is the most efficient;
  - Properly using the prediction results can increase the gross profit per room by around 47%.
- What can we do further?
  - Model: Add features regarding different hotels, countries, etc.
  - Cost matrix: Take re-booked rooms after cancelled into account.



THANKS!

Questions?