



## Predicting Hotel Cancellations

► Kilmar Lazo



# Contents

1

Business  
Understanding

2

Variables Used

3

Data  
Understanding

4

Modeling and  
Evaluation

5

Conclusion

# Business Understanding



## **Business Stakeholder**

- Hotels

## **Business Problem**

- Predict hotel cancellations to increase revenue and operate efficiently

## **Benefits**

- Improved efficiency
- Increased revenue
- Enhanced customer satisfaction
- Reduced costs
- Improved decision-making

# Variables Used

---

## Is Canceled

### Target Variable

- Value indicating if the booking was canceled (1) or not (0)

---

## Month

### Variable Used

- Month of arrival date

---

## Deposit Type

### Variable Used

- Type of deposit customer made to guarantee the booking.

---

## Lead Time

### Variable Used

- Number of days between the booking date and the arrival date

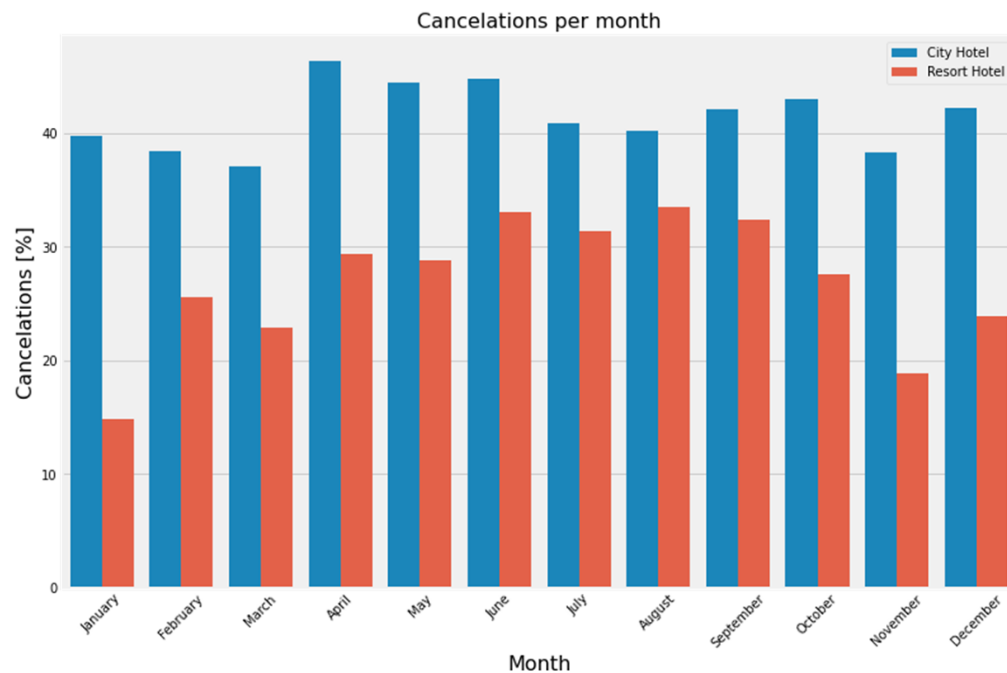
---

## Is Repeated Guest

### Variable Used

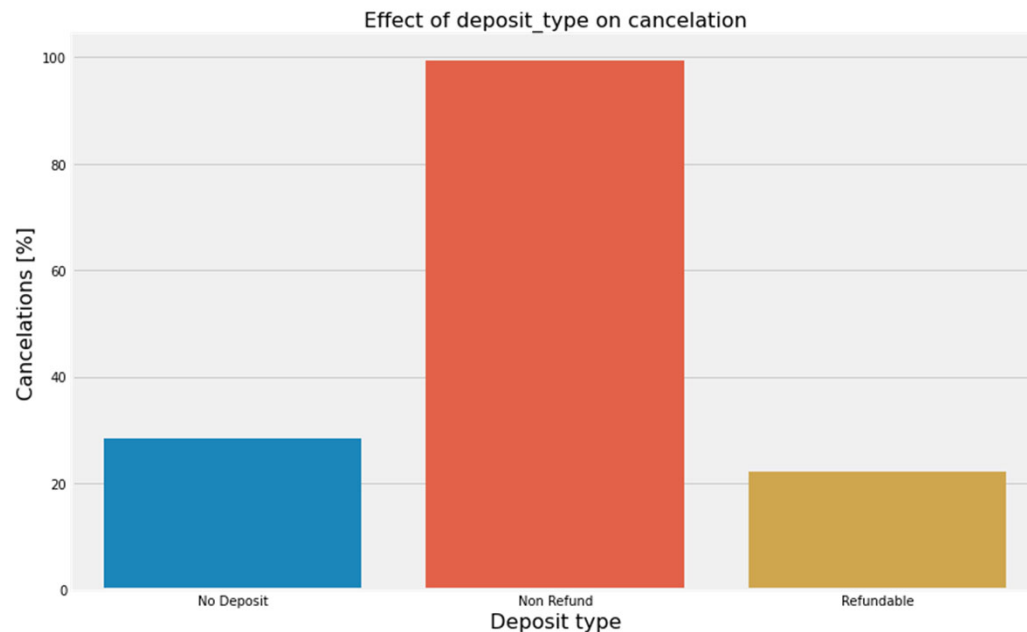
- Value indicating if the booking was made by a repeated guest (1) or not (0)

# Data Understanding



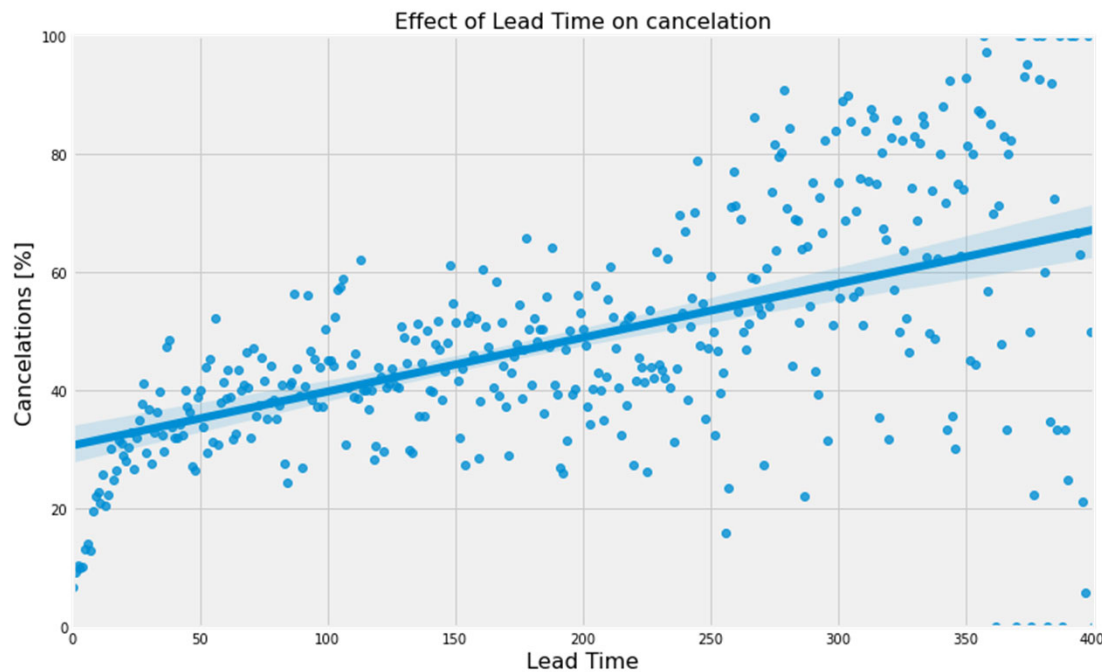
- City Hotel's cancellations were consistent throughout the year
- Resort Hotel faced higher cancellation percentages during the Summer months

# Data Understanding cont.



- Bookings with a non-refundable deposit were almost always cancelled
- Bookings with refundable deposits were least likely to get cancelled

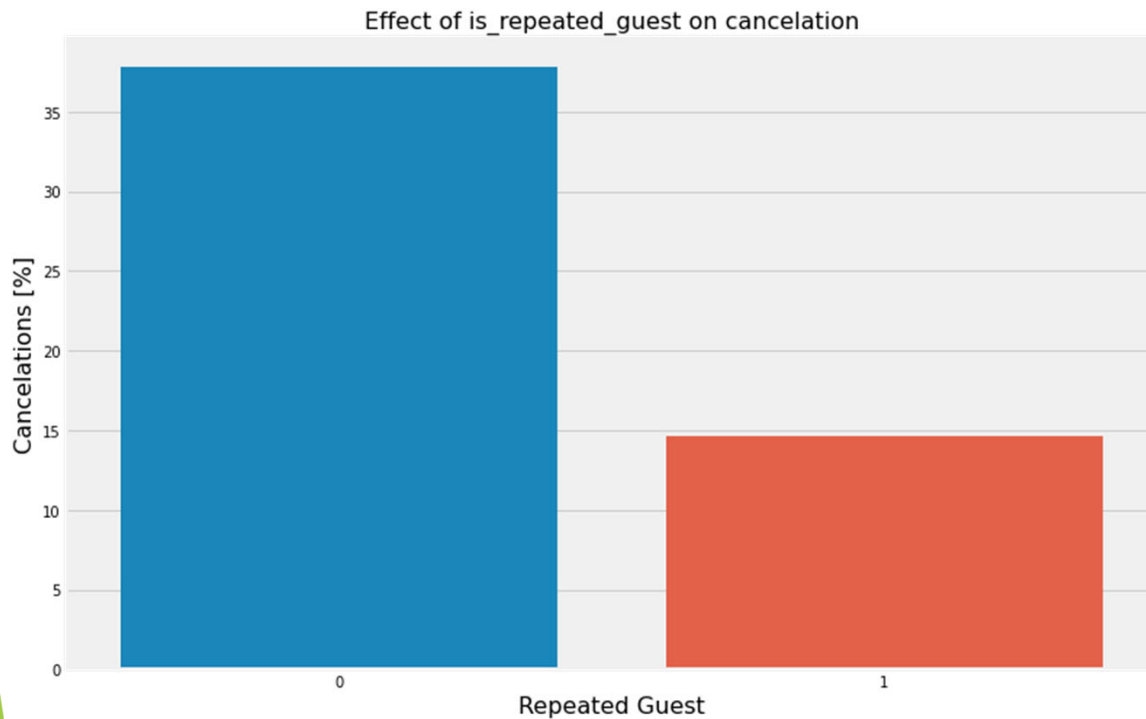
# Data Understanding cont.



Copyright © 2020 Deloitte Development LLC. All rights reserved.

- Lead time has a strong positive correlation with our target variable (is canceled)
- As lead time increases, the chance of cancellation increases

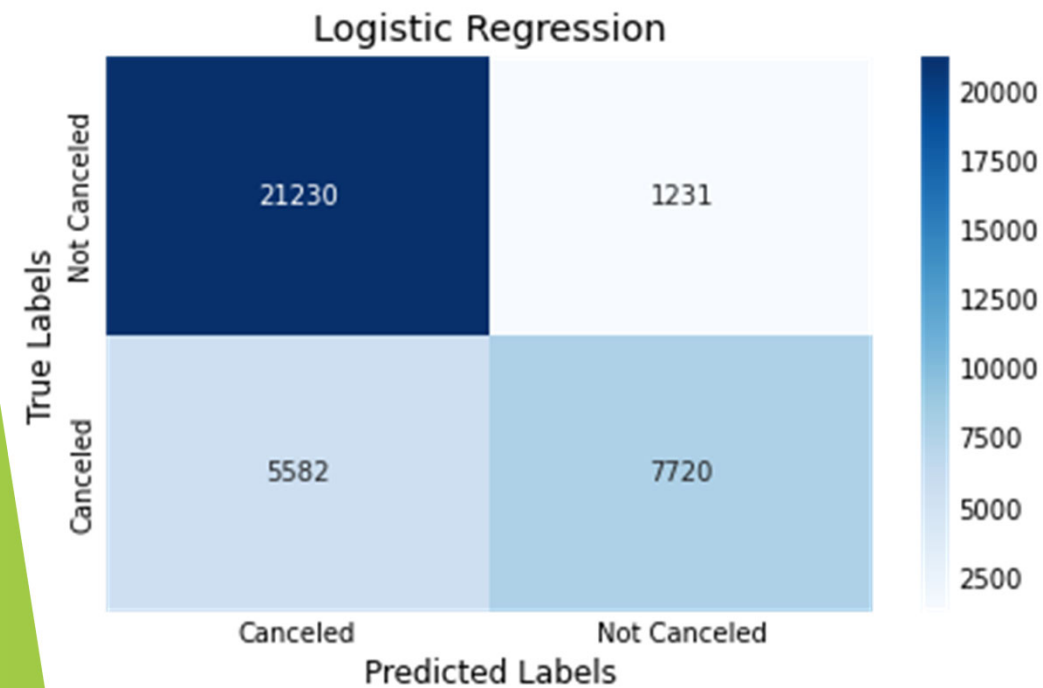
# Data Understanding cont.



- Guests that have never been to the hotel are more likely to cancel their booking
- Returning guests are less likely to cancel their booking

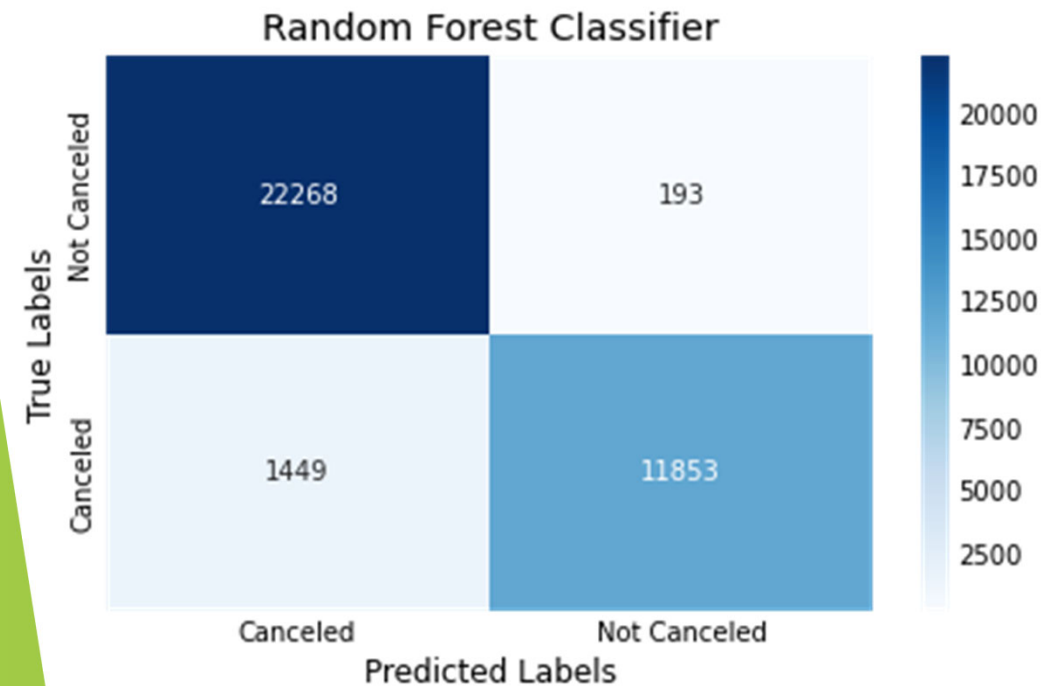


# Model Evaluation - Logistic Regression



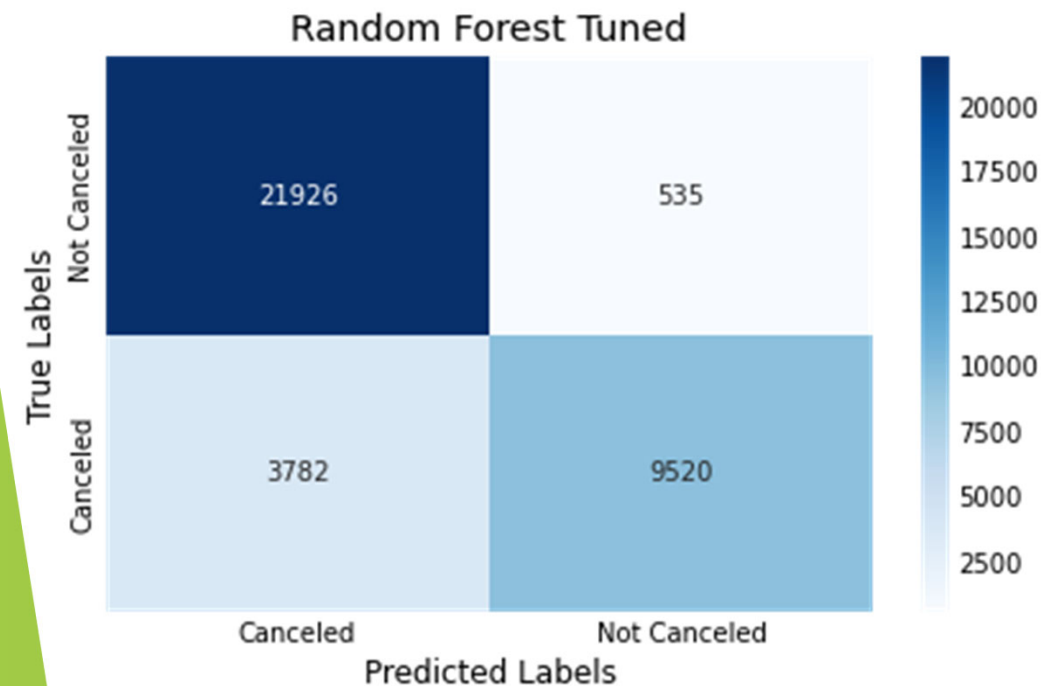
Accuracy	81%
Recall Score	95%

# Model Evaluation - Random Forest Classifier



Accuracy	95%
Recall Score	99%

# Model Evaluation - Random Forest Classifier Tuned



Accuracy	87%
Recall Score	99%

# Conclusion



## Final Model

- Random Forest Classifier with an accuracy of 95% and a recall score of 99%

## Limitations

- There may be more factors that affect cancellation
  - Hotel's location, number of hotels around the area, hotel rating

## Next Steps for Improvements

- Gather more samples with extra columns
- Gather more samples from hotels across the major metropolitan areas, along with smaller metropolitan areas