Deloitte.

Predicting Hotel Cancellations

▶ **Kilmar Lazo**

MAKING AN
IMPACT THAT
MATTERS
since 1845

Contents

1	2	3	4	5
Business Understanding	Variables Used	Data Understanding	Modeling and Evaluation	Conclusion

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.

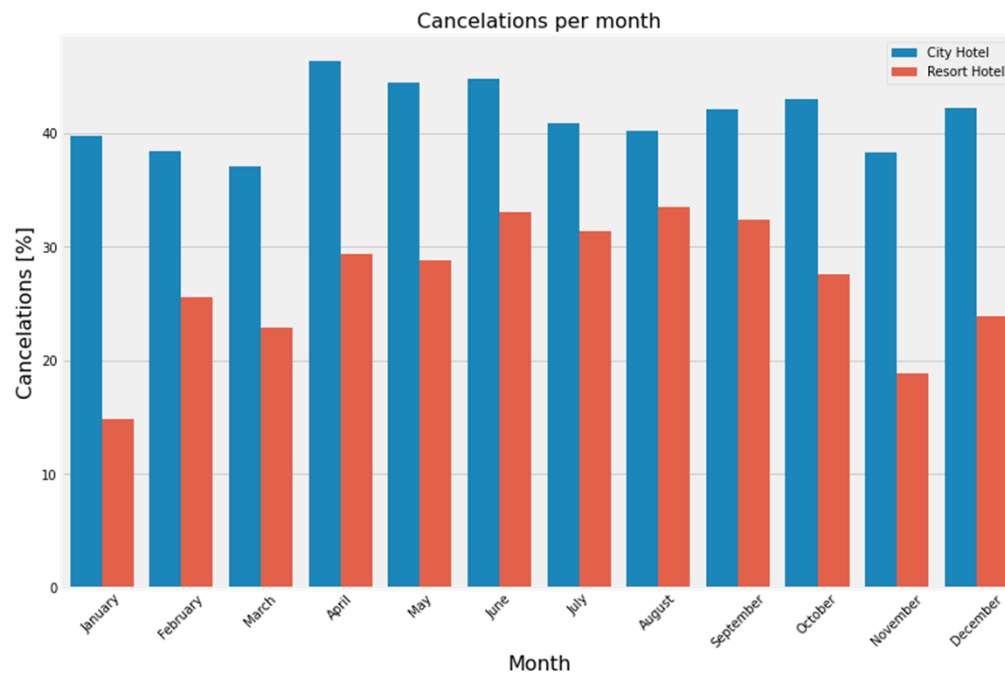## Lead Time

Variable Used

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

Is Repeated Guest
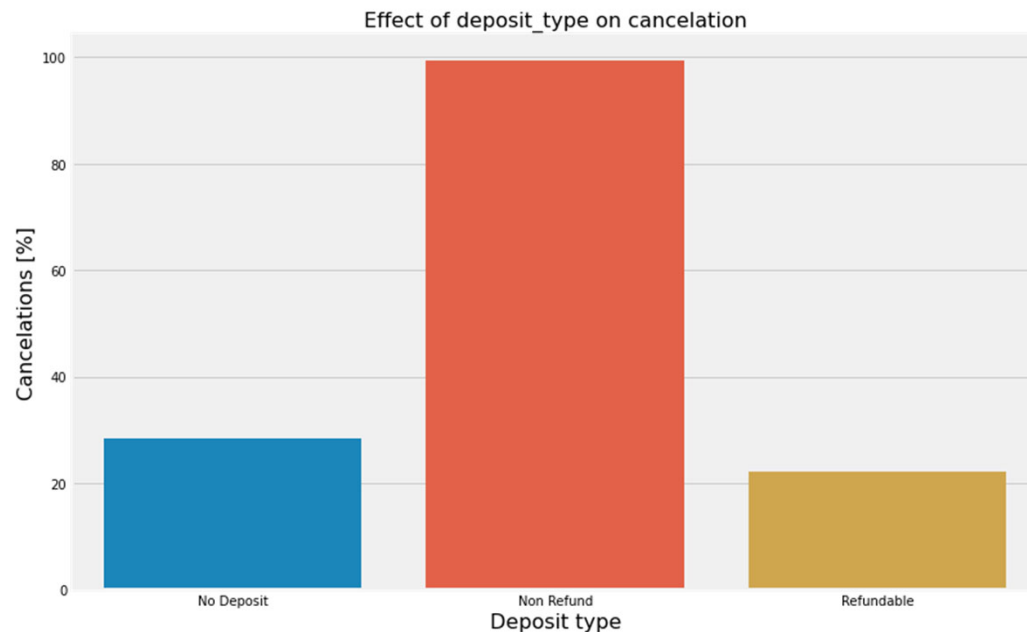
Variable Used

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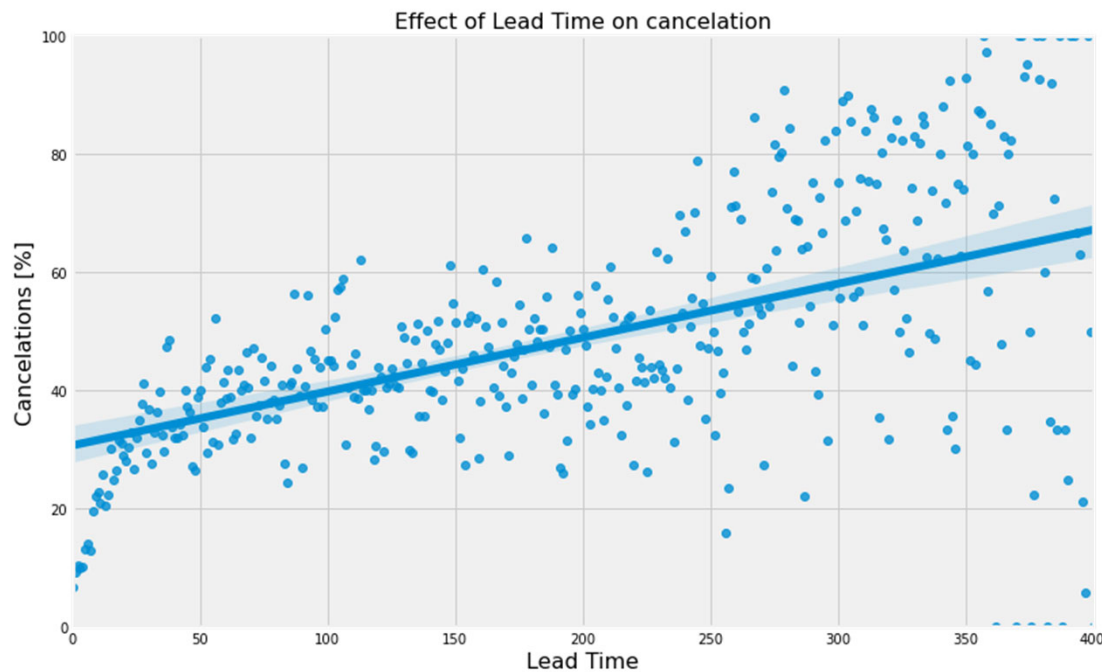
Data Understanding

Cancelations per month

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

Data Understanding cont.

Effect of deposit_type on cancelation

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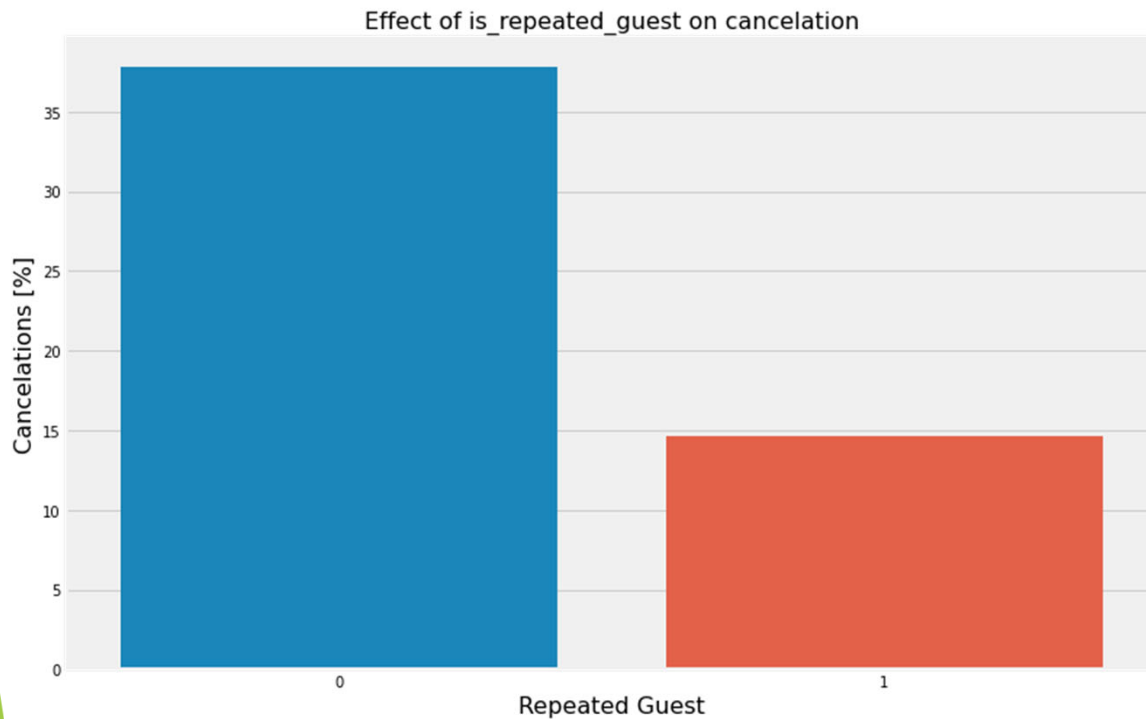- Bookings with refundable deposits were least likely to get cancelled

Data Understanding cont.

Effect of Lead Time on cancelation

- Lead time has a strong positive correlation with our target variable (is canceled)
- As lead time increases, the chance of cancellation increases
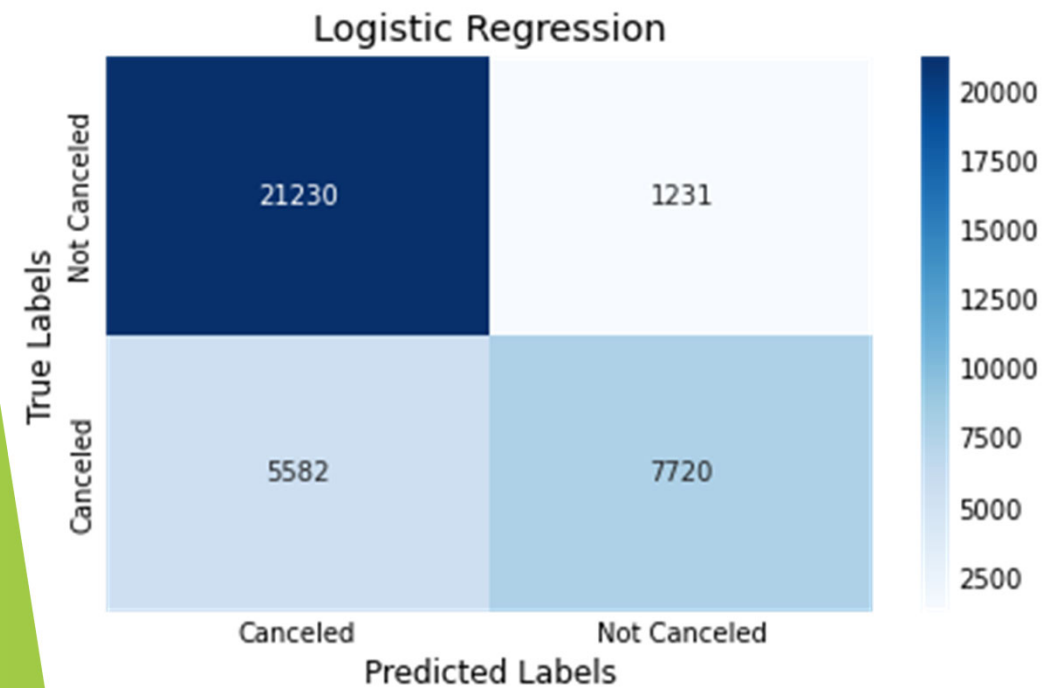
Data Understanding cont.
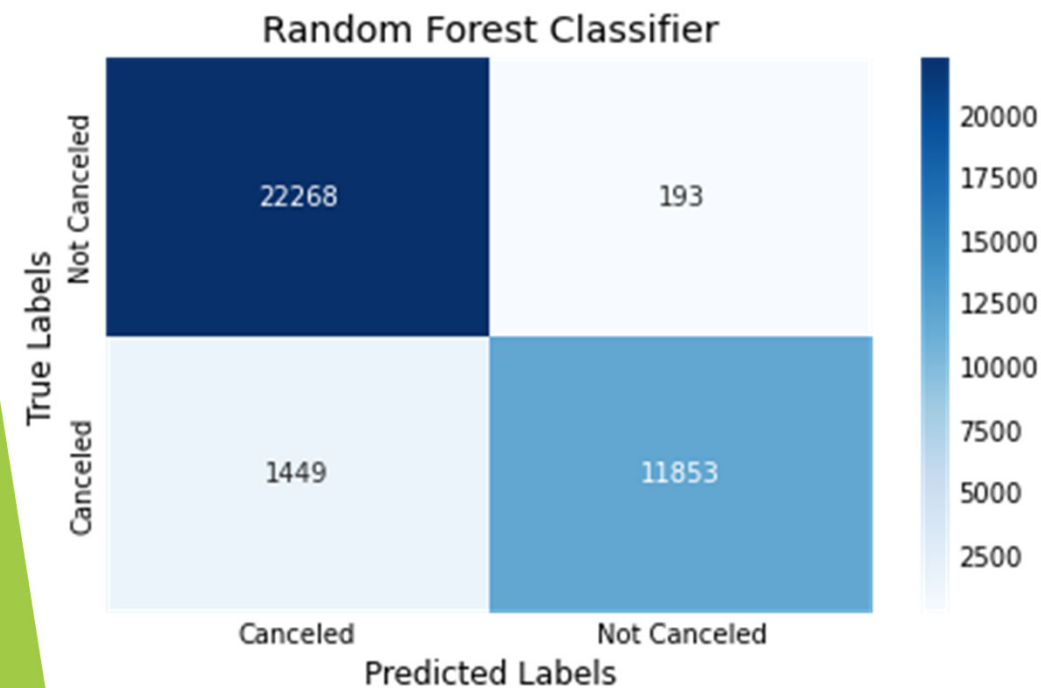
Effect of is_repeated_guest on cancelation

- 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

Random Forest Classifier

	Predicted: Canceled	Predicted: Not Canceled
True: Not Canceled	22268	193
True: Canceled	1449	11853

True Labels / Predicted Labels

Colorbar scale: 20000, 17500, 15000, 12500, 10000, 7500, 5000, 2500

Accuracy	95%
Recall Score	99%

Model Evaluation – Random Forest Classifier Tuned

Random Forest Tuned

	Predicted: Canceled	Predicted: Not Canceled
True: Not Canceled	21926	535
True: Canceled	3782	9520
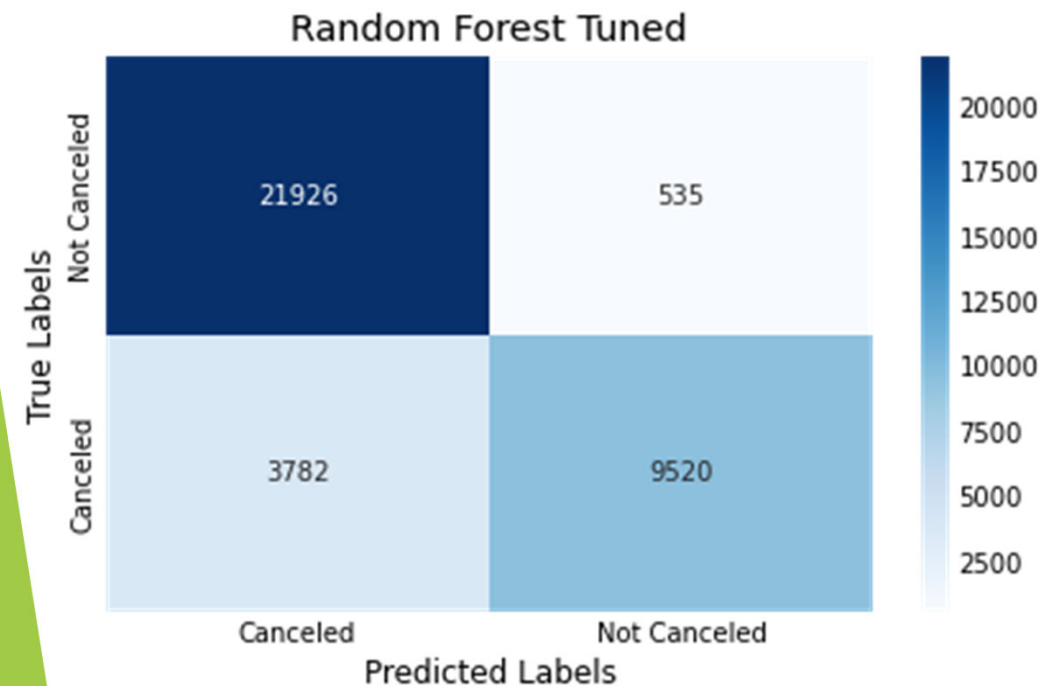
Accuracy	87%
Recall Score	99%

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