



Cancel Less, Live More: An Analysis on Hotel Booking Cancellations

Edith Magana, Morgan Allen, Woo Seok Kim, Marisol Mondragon

Overview

- Description of data
- Analysis
- Model (XGBoost)
- Suggestions

Description of Data

Data

From article Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes

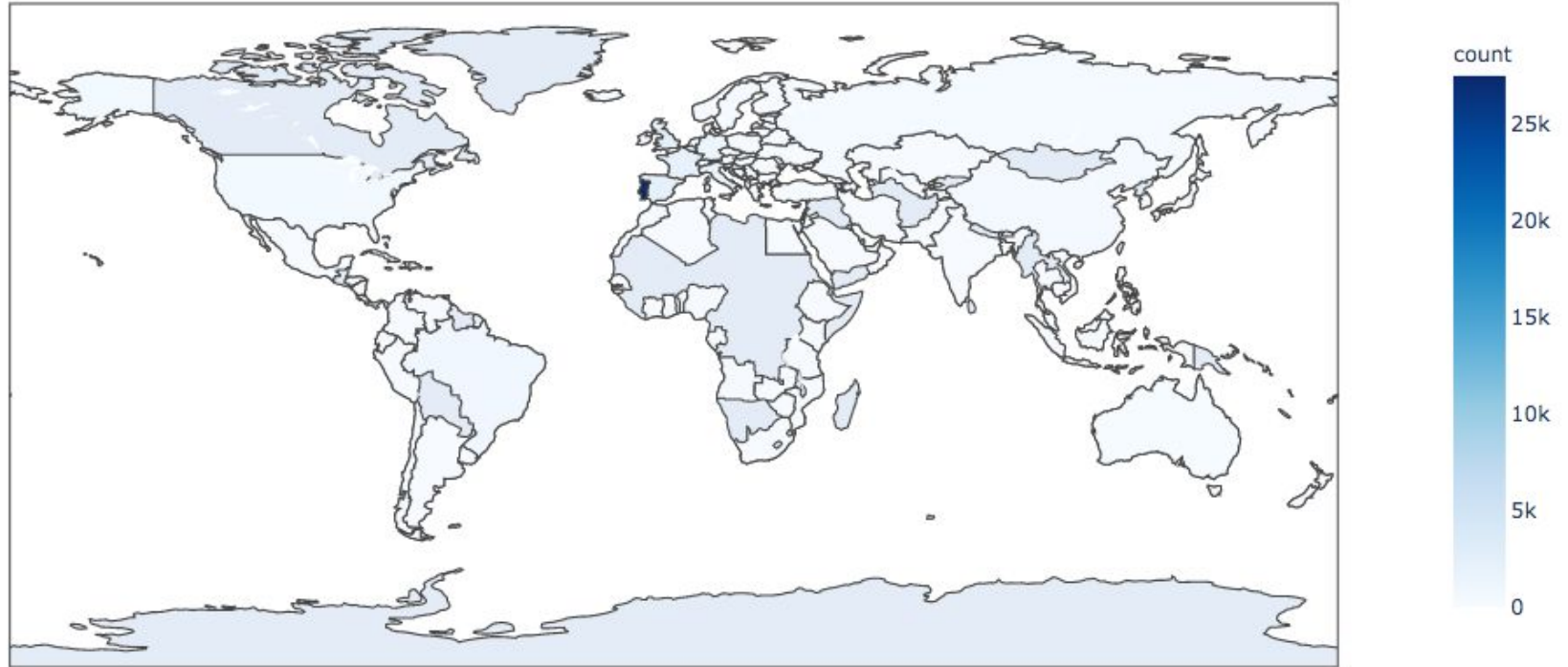
Rows - 118,898

Columns - 31

- Features
 - booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, etc.

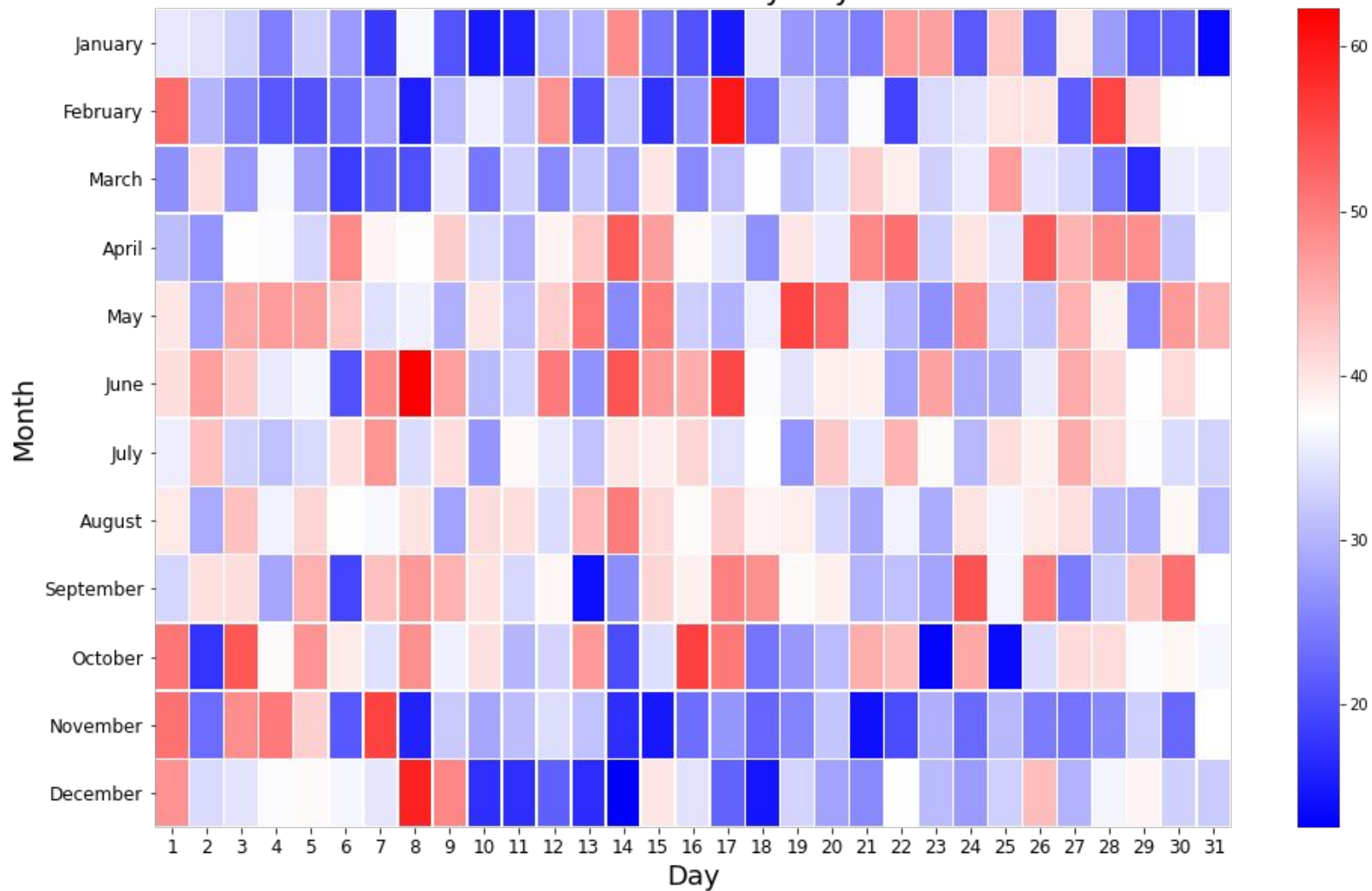
Analysis

Home Country Of Reserves



Features	Canceled	Not Canceled
Lead Time	144.93	80.32
Repeated Guest	1.30%	4.40%
Previous Cancellations	0.208	0.016
Previous Bookings w/o Cancellation	0.024	0.195
Booking Changes	0.098	0.294
Days In Waiting List	3.57	1.59
Average Daily Rate	105	100
Car Parking Space	0	0.1
Special Requests	0.329	0.715
From Portugal	61%	28%
Non Refund	33%	0.1%

Cancellations by Day



June 8th

- This is considered a high risk for cancellation day since over 40% of hotel bookings were cancelled.
- June 10th is a Portugal Day (National Holiday)
- Assume most people are coming in to town for that day on June 8th

Features	June 8
Lead Time	105.46
Repeated Guest	2.7%
Previous Cancellations	0.003
Previous Bookings w/o Cancellation	0.059
Booking Changes	0.104
Days In Waiting List	0.10
Average Daily Rate	129
Car Parking Space	0.02
Special Requests	0.373
From Portugal	54%
Non-Refund	29%



December 14

- Nothing significant about December 14th in Portugal
- Not considered high risk since cancellations were under 40%

Features	December 14
Lead Time	22.29
Repeated Guest	11.6%
Previous Cancellations	0.071
Previous Bookings w/o Cancellation	0.964
Booking Changes	0.241
Days In Waiting List	0.00
Average Daily Rate	57
Car Parking Space	0.11
Special Requests	0.625
From Portugal	53%
Non Refund	0%



Model (XGBoost)

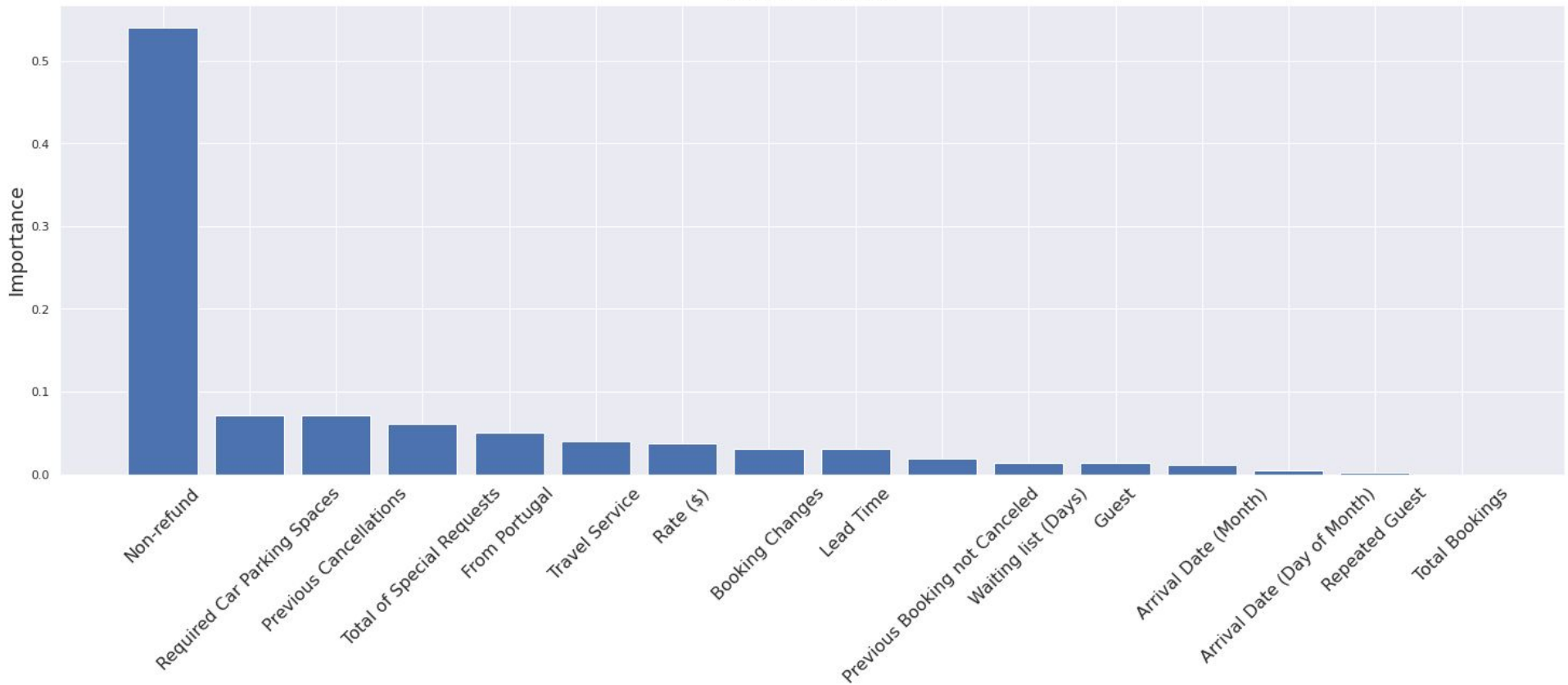
XGBoost

- Decision tree-based model
- Decision trees can capture non-linearity in data
- In XGBoost, multiple decision trees are built subsequently (called “boosting”)
- Builds current decision tree based on where previous decision tree did poorly

XGBoost

- Used the columns discussed in the previous tables
- Model will help Hotels reduce cancellations on hotel bookings
- 82% precision: out of the customer we predicted will cancel, 82% of them actually did
- Predicting cancellations well

Model Features



Suggestions

Lead Time

- The amount of time between booking and the actual reservation
- The longer the lead time the more likely you are to cancel

Suggestion:

- If someone books far in advance constant contact needs to take place between the customer and hotel to keep them interested
- Possibly offer deals halfway in their lead time

Previous Bookings

- Using Customer accounts, look back at the previous cancellations and previous bookings without cancellation

Suggestion:

- People with a lot of bookings without cancellation should be categorized as low risk
- People with a lot of bookings with cancellation should be high risk and should be offered deals and/or should be contact more often

Days In Waiting List

Suggestion:

- If someone is in the waiting list for 2+ days, they should be given a deal or discount on their booking to encourage them not to cancel

From Country of the Hotel

- Whether they are from the country the hotel is in (in our case Portugal) or not
- If they are from out of the country, they are less likely to cancel

Suggestion:

- People from in the country should get discounts or rewards
- Also keeping in constant contact before arrival with someone booking in country could help them not want to cancel

Non Refundable Booking

- Bookings can either be No Deposit, Non Refundable, or Refundable
- The majority are No Deposit but the Non Refundable bookings seem to cancel 99% of the time.

Suggestion:

- Do not offer Non Refundable Deposit Hotel Bookings
- If you do offer them, give a discount and stay in contact with the person booking the hotel to help them not cancel

Conclusion

- Keep in contact with your customers
- Give them discounts
- Make sure to have good reviews
- Give them the best services to all clients



Thank you

Analysis

Feature	June 8	Canceled	Not Canceled
Lead Time	105.46	144.93	80.32
Repeated Guest	2.7%	1.30%	4.40%
Previous Cancellations	0.003	0.208	0.016
Previous Bookings w/o Cancellation	0.059	0.024	0.195
Booking Changes	0.104	0.098	0.294
Days In Waiting List	0.10	3.57	1.59
Average Daily Rate	129	105	100
Car Parking Space	0.02	0	0.1
Special Requests	0.373	0.329	0.715
From Portugal	54%	61%	28%
Non Refund	20%	22%	0.1%

Analysis

Feature	December 14	Canceled	Not Canceled
Lead Time	22.29	144.93	80.32
Repeated Guest	11.6%	1.30%	4.40%
Previous Cancellations	0.071	0.208	0.016
Previous Bookings w/o Cancellation	0.964	0.024	0.195
Booking Changes	0.241	0.098	0.294
Days In Waiting List	0.00	3.57	1.59
Average Daily Rate	57	105	100
Car Parking Space	0.11	0	0.1
Special Requests	0.625	0.329	0.715
From Portugal	53%	61%	28%
Non Refund	0%	33%	0.1%

	precision	recall	f1-score	support
0	0.81	0.92	0.86	18751
1	0.82	0.62	0.71	10974
accuracy			0.81	29725
macro avg	0.81	0.77	0.78	29725
weighted avg	0.81	0.81	0.80	29725

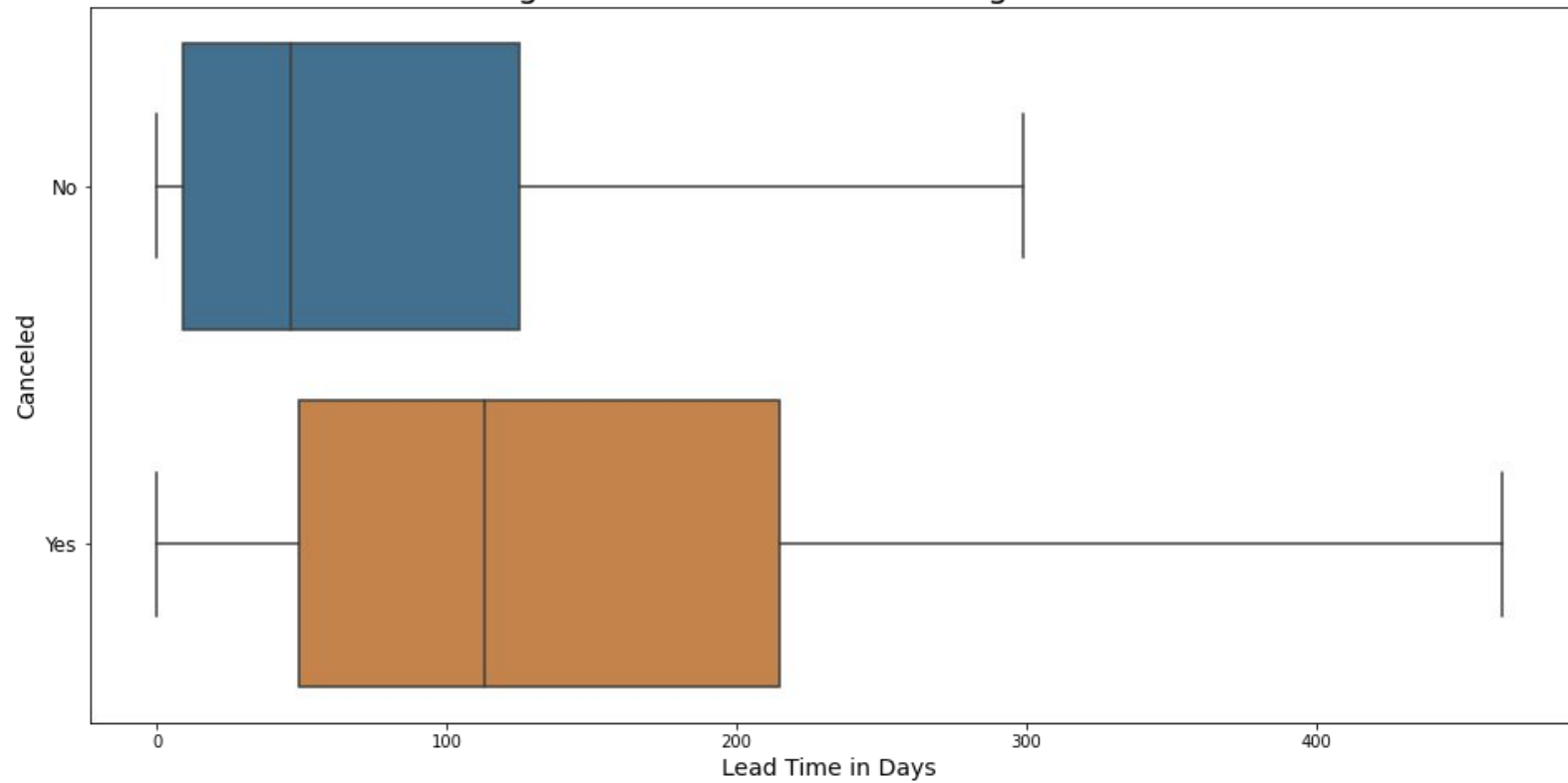
For 0 (did not cancel):

- 81% precision: out of the customers we predicted will not cancel, 81% of them actually did not
- 92% recall: out of the customers who actually did not cancel, 92% of them did we predict would not cancel
- 86% f1-score: the harmonic mean rather than the normal mean of precision and recall for not cancel

For 1 (cancel):

- 82% precision: out of the customer we predicted will cancel, 82% of them actually did
- 62% recall: out of the customers who actually did cancel, 62% of them did we predict would cancel
- 71% f1-score: the harmonic mean rather than the normal mean of precision and recall for cancel

Length of Time Between Booking and Arrival



Total Special Requests

- The total amount of extra things someone asks for (extra towels, early checkout, etc.)

Suggestion:

- If someone does not have a special request, ask them multiple times if they need anything extra
- If someone requests more, they will be more likely to not cancel

Average Daily Rate



Suggestion:

