Identifying the flake!

A classification model to identify the potential cancelling customers for Marriott International

Marriott INTERNATIONAL

Milad Afrasiabi

October 2021

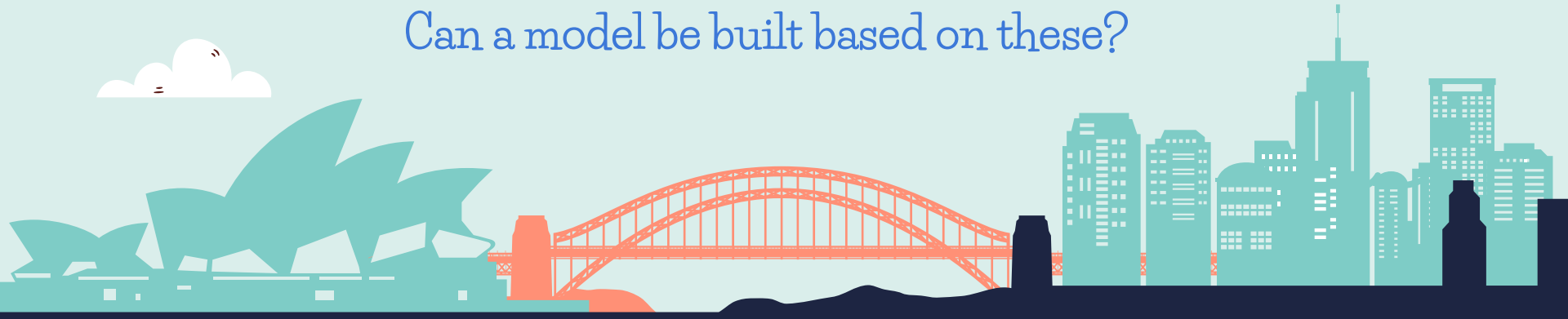
Overview and Problem

- COVID forced for laxer cancellation policies
- Cancellation are costly to a hotel business
- Non-systematic approach is inefficient

Question

Are there variables that can accurately predict if a given reservation prone to cancellation?

Can a model be built based on these?



Beneficiaries

01 Marriott International

Marriott INTERNATIONAL

02 Travel agents working with them

Impact

Reduce Cancellations
&
Maximizing Profit

Identifying cancellation-prone bookings

Adopting proper strategy (harsher policies, promotional offers, etc.)

Data Science Solution

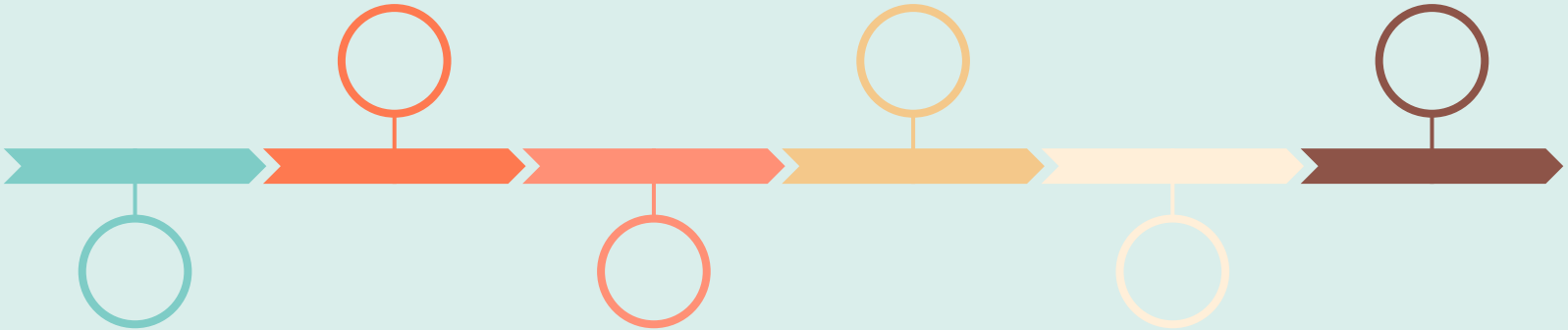
- Building a classification model to identify cancellation-prone bookings

- Using these models to identify the key variables and act on them

Critical Assumption

A single booking features can be accurate predictors of its chance of cancellation

Data



Booking Data

2 Hotels in Portugal

Source

Antonio et al
Data in Brief, Feb 2019

31

of original features

119390

of observations

2015 to 2017

Recording period

2 Target Classes

Canceled
Not Canceled

Features

- Average Daily Rate
- Lead Time
- Deposit Type
- Month of Arrival
- Meal Type
- Total Stay Nights

Methodology

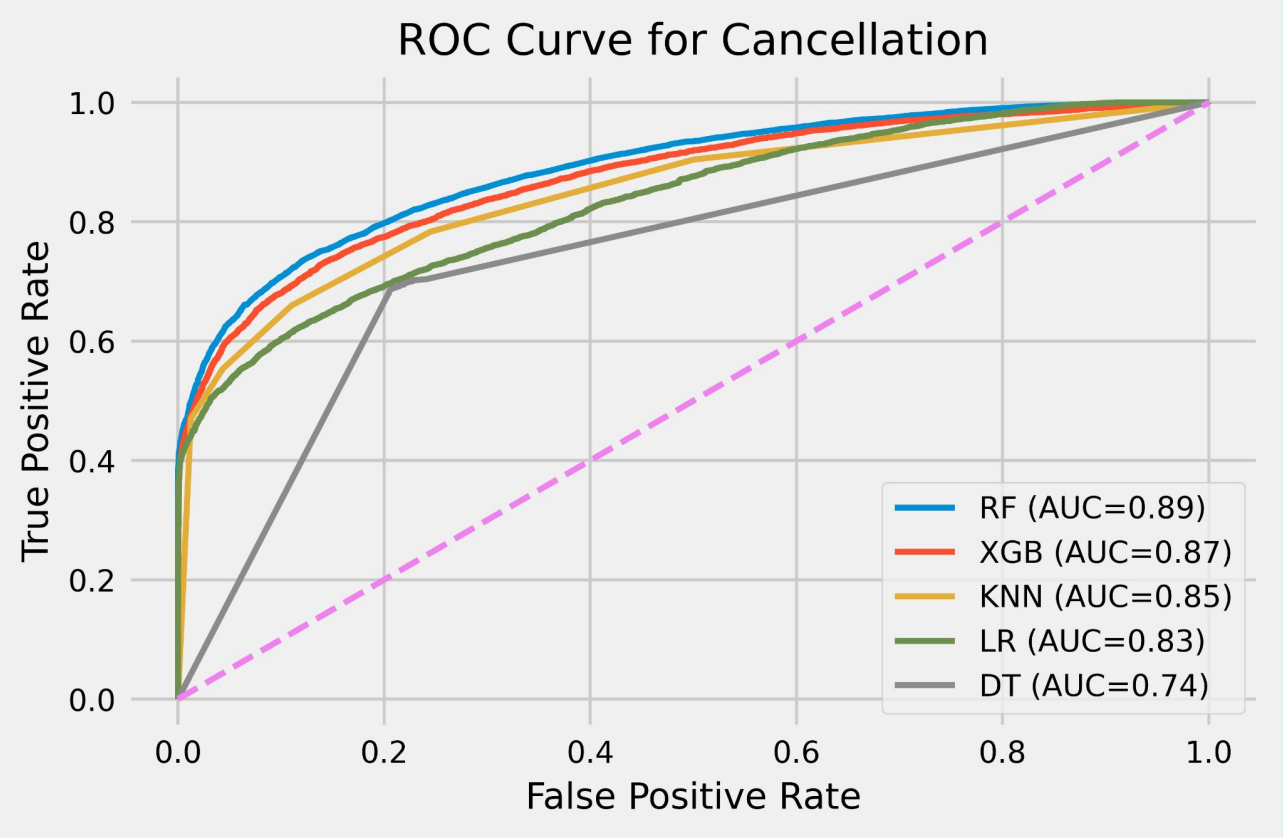
01 Train-Test Split

02 Feature Addition & Engineering

03 Model Fitting
KNN
Logistic R
Decision Trees
Random Forest
XGBoost

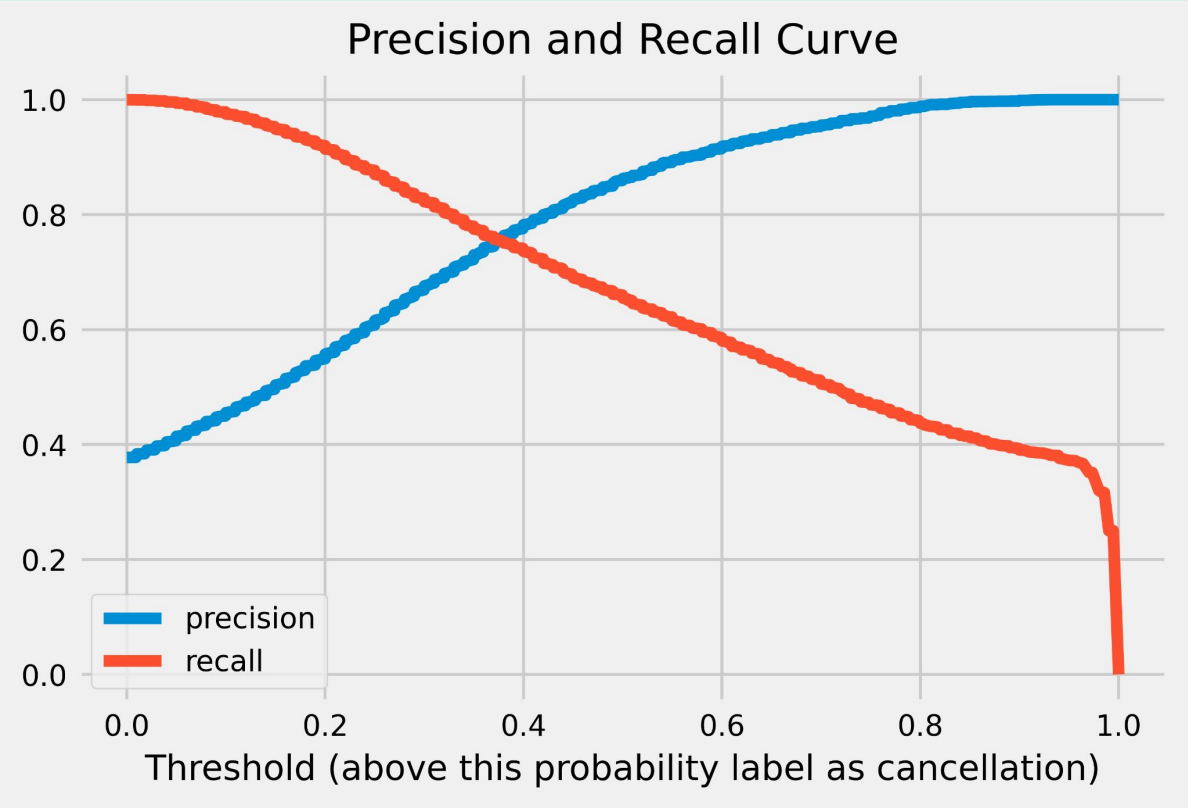
04 Model Selection & Tuning

Results



Random Forest

Precision: 0.834
Recall : 0.7
AUC : 0.9



Threshold:

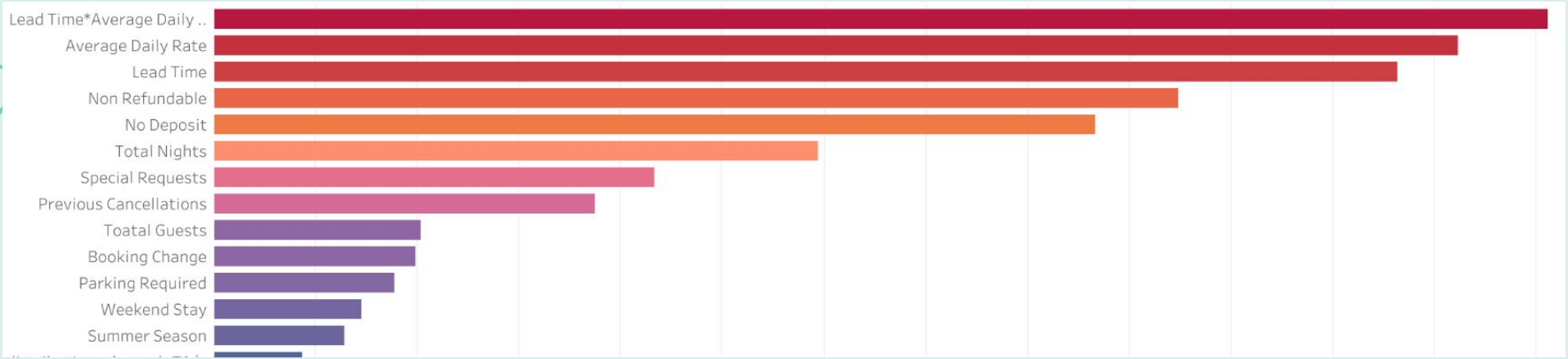
0.3775

Final Model Performance

Random Forest
(Threshold =0.3775)

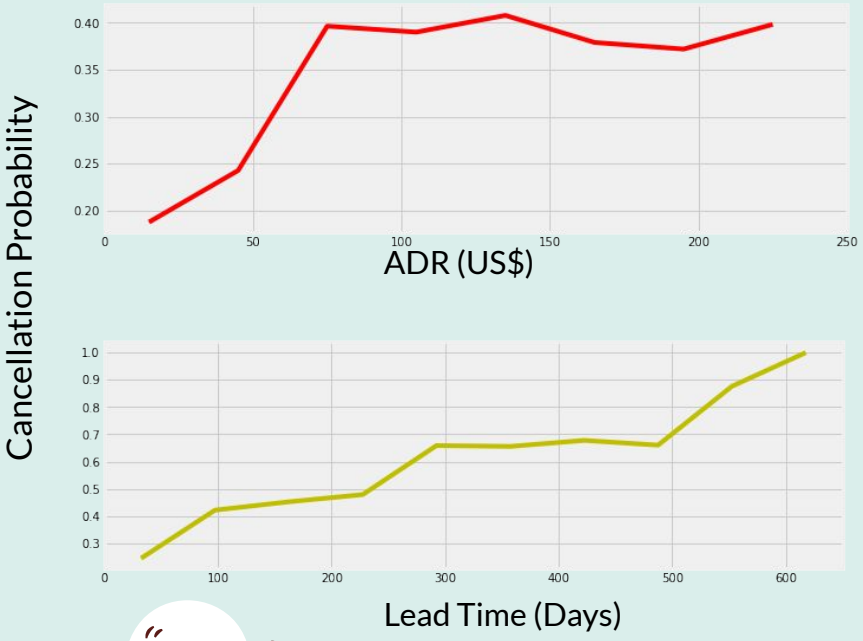
- Accuracy = 0.821
- Precision = 0.751
- Recall = 0.788
- F1 = 0.769
- ROC AUC = 0.9

Feature Importance



- Lead Time
- Average Daily Rate
- Deposit Type

Features


Cancellation Probability
0.40
0.35
0.30
0.25
0.20
0
50
100
150
200
250
ADR (US$)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
100
200
300
400
500
600
Lead Time (Days)


1.0
0.8
0.6
0.4
0.2
0.0
No Deposit
Full Deposit
Partial Deposit

Takeaways

- Hotel cancellation can be accurately predicted buy a handful of relevant variables
- The Random Forest model built in this project does this task with a relatively high performance
- Average Daily Rate, Deposit Type and Reservation Lead Time came out as the most relevant features out of this model

Future Work

Further Feature Engineering

XGBoost Model fine tuning

Thanks!

Questions?