HOTEL CANCELLATIONS PREDICTION

A Revenue Opportunity

By Reji Oomman ISP Jul 2024

Executive Summary

Nearly 40% online bookings are cancelled prior to arrival according to study done by D-edge hospitality solutions. The reports states that guests have become accustomed to free cancellation policies that have been made popular by online travel agencies. The aim here is to correctly predict potential cancellation using a machine learning model.

Exploratory Analysis

Finding patterns and trends with the help of given features along cleaning and preprocessing the data for machine learning models.

Model Building

Different Classification model build and Final model optimized with the help of gradient boosting classifier

Inferences

After model prediction, we were able to define 5 business recommendations with the help of model prediction results.

Study figures

Business Objective

Our aim is to find patterns in hotel reservations for two different hotel types and predict potential hotel cancellations with the help of a machine learning model.

By Reji Oomman ISP Jul 2024

FEATURES

Exploratory Analysis

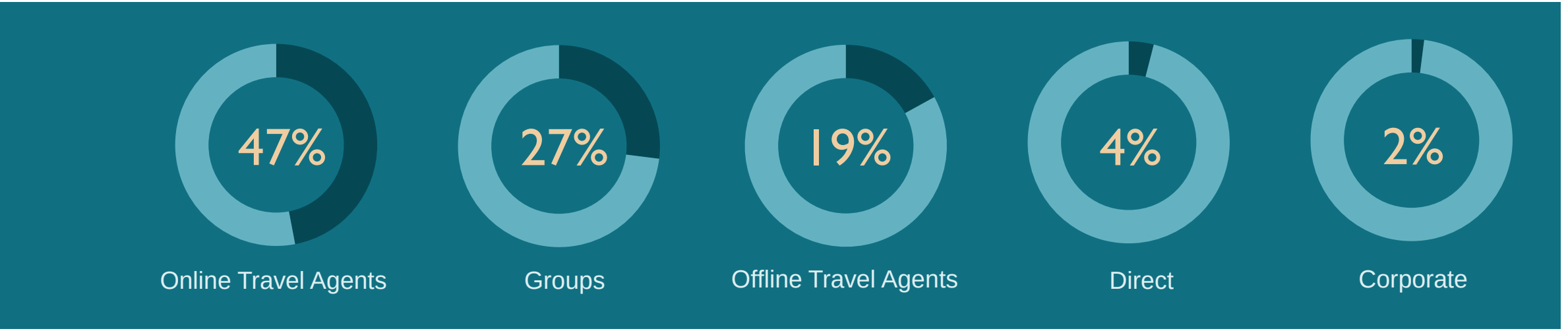
28	66%	77%	41%	47%	88%
Features	*Reservations are for city hotel*	*Reserved guests prefered room with only breakfast meal plan*	*Guests are from Portugal among 177 nationalities*	*Reservations are from OTA market segment*	*Reservations are having no deposit as guarantee.*

37%

Records are of cancelled reservations which is slightly higher than that of industry standards of 28%.

By Reji Oomman ISP Jul 2024

MARKET SEGMENTS: Cancellation insights



Market Segments	Type_1	Type_2	Overall
OTA*	32.89%	14.18%	**47.07%**
Groups	21.83%	5.62%	**27.45%**
Offline TA**	16.26%	2.58%	**18.84%**
Direct	2.40%	2.00%	**4.39%**
Corporate	1.45%	0.80%	**2.25%**
Total	**75%**	**25%**	**100%**

Maximum cancellations are from OTA segments and in city hotel.

** OTA- Online Travel agents*

*** TA – Travel agents*

Model Building: Preprocessing

Missing Values

Missing values in three columns are treated using simple imputer package with 'most frequent' strategy.

Duplicate records

33964 Duplicate records were found. Almost 28% of the records had to be dropped for ML models

Outlier Treatments

Numerical columns had outliers which are capped with lower and upper limit using inter quartile range.

Scaling

Scaling was required for numerical features to bring the features in similar scale.

Dummy Encoding

Categorical features required dummy encoding for machine learning model (ensured drop first option is used to avoid multicollinearity).

Train / Test Split

The cleaned data is further split into two parts, Train and Test. Train data is 70% and Test data is 30% along with stratification method.

FINAL MODEL : Gradient Boosting Classifier

After trying 9 different Machine learning models, Final model had been gradient boosting decision tree classifier.

Key Parameters

- *Base estimator:* ***Decision Tree***
- *Learning rate:* ***0.1***
- *No. of estimators:* ***100***
- *Min sample leaf:* ***1***
- *Min sample split:* ***2***

Model Metrics

0 – Not cancelled 1 - Cancelled	Train		Test	
	0	1	0	1
Precision	0.84	0.68	0.84	0.67
Recall	0.91	0.53	0.91	0.52
F1-score	0.88	0.59	0.87	0.59
Accuracy	0.81		0.81	
AUC score	0.86		0.86	

As both precision and recall are essential therefore F1-score was used as most important metric for choosing the optimum and stable model.

By Reji Oomman ISP Jul 2024

Business Recommendations

Revenue opportunity

Based on model classification we can identify the potential cancellations in advance and overbook certain dates to avoid unsold rooms in eleventh hour.

Revenue per available room

If the cancellations are recognized well in advance then room rates can be adjusted accordingly to avoid last minute discounts.

Segment allocation

Cancellations known earlier can help in allocation of inventory of rooms to different room segments.

Resource planning

Potential cancellations acknowledged by the model can assist in resource planning in hotels such as staff rostering, menu planning and preventive maintenance.

Brand reputation

Cancellations known earlier can help in keeping tap on overbooking limits and prevent guest complaints from turn away situations.

By Reji Oomman ISP Jul 2024

Challenges & Opportunities

Data Constraints

Data was little out of date the trends would have changed over the time.

Local Resources

The Machine learning models were crashing often due to limited local computing resources.

HPC Computation

Need to try for HPC computation and analysing through DASK dataframe

Future Opportunities

Data recency can identify the following:

- Overbooking Trends
- Market segmentation trends
- Revenue Management

By Reji Oomman ISP Jul 2024

THANK YOU!
Q&A

Reji Oomman

ISP July 2024