

Hotel Booking EDA + Cancellation prediction

Introduction

his data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things

Dataset

From Kaggle it have 119390 entries and 32 columns

Data Cleaning

Drop duplicate values

replace all null values with 0 and unknow for country column

Drop company column because has many null values

Exploratory Data Analysis (EDA)

From where the most guests are coming

Which are the most busy months

How long do people stay at the hotels

Modeling

0.80, Logistic Regression

0.85 , K Nearest Neighbors (KNN)

0.92 , Decision Tree Classifier

Conclusion

By these models the hotel managers enables to decrease the loss from booking cancellations and to decrease the risks associated with overbooking