

HOTEL BOOKING PREDICTION

By: Putu Rika Sahriana

...



Table of Contents



BACKGROUND



OBJECTIVE



RESULT SUMMARY

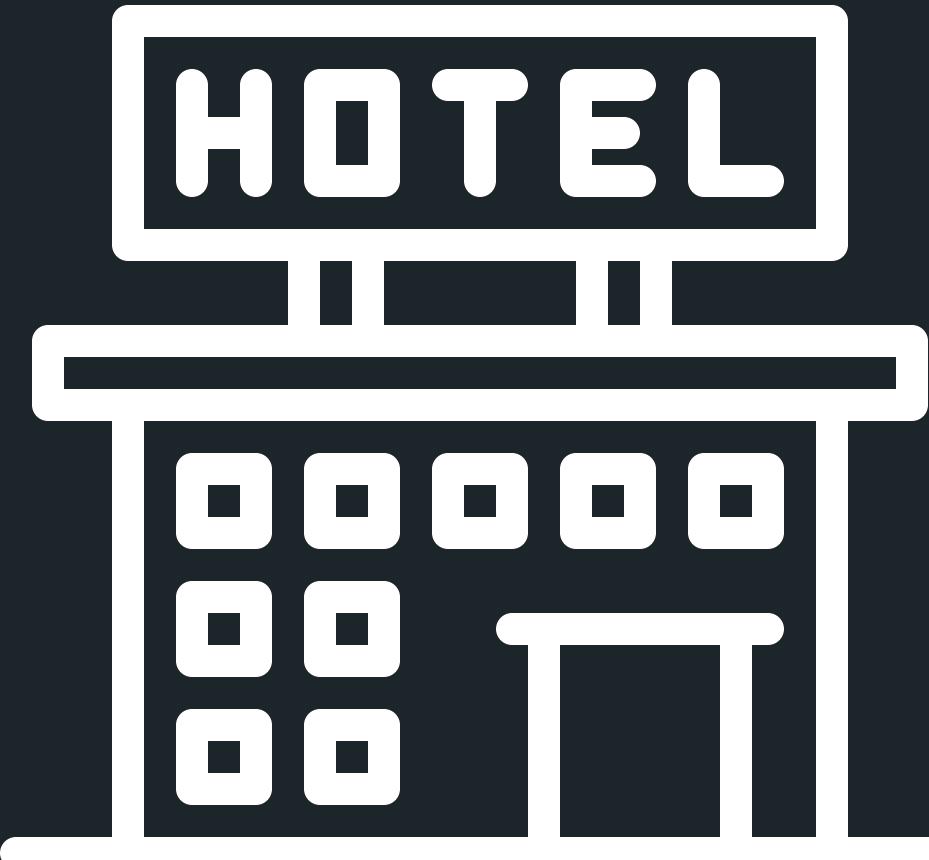
“

BACKGROUND

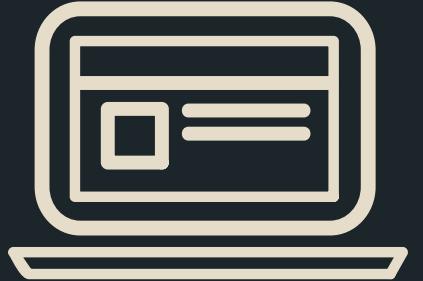
Hotel is a business that depends on the number of visitors. In order to get maximum profit is requires a lot of visitors.

Hotels also have a certain capacity, so there are limitations in accepting visitors.

Because of these problems, the hotel provides a booking system, so that visitors can order services in advance and get services according to the desired day.



PROBLEM



WHAT CAN WE DO TO MINIMIZE LOSSES?

The solution to minimize losses due to booking cancellations is to look at customer behavior which can be taken into consideration when making decisions.



BACKGROUND

ABOUT DATA

To provide a solution to this problem, I did some analysis using the hotel_booking dataset from kaggle with 119.390 rows and 32 columns



The features contained in this dataset:

- HOTEL
- IS CANCELED
- LEAD TIME
- ARRIVAL DATE YEAR
- ARRIVAL DATE MONTH
- ARRIVAL DATE WEEK NUMBER
- ARRIVAL DATE DAY OF MONTH
- STAYS IN WEEKEND NIGHTS

- STAYS IN WEEK NIGHTS
- ADULTS
- CHILDREN
- BABIES
- MEAL
- COUNTRY
- MARKET SEGMENT
- DISTRIBUTION CHANNEL

- IS REPEATED GUEST
- PREVIOUS CANCELLATIONS
- PREVIOUS BOOKINGS NOT CANCELED
- RESERVED ROOM TYPE
- ASSIGNED ROOM TYPE
- BOOKING CHANGES
- DEPOSIT TYPE
- AGENT

- COMPANY
- DAYS IN WAITING LIST
- CUSTOMER TYPE
- ADR
- REQUIRED CAR PARKING SPACES
- RESERVATION STATUS
- RESERVATION STATUS DATE

Preparing Data for Analysis and Modeling

01



02



03



04



Check Missing Value

Country (488), Agent (16340) and Company (112593)

Fill NA

I fill missing value with 0 to simplify data processing

Drop Columns

Some value it doesn't make sense is dropped. There are columns adult, children, babies that have age with 0 value (180)

Final Data

There are 32 columns with 119.210 rows



Business Question



Where do the guests come from?



How much do guests pay for a room per night?



How does the price per night vary over the year?



Which are the most busy month or in which months Guests are high?



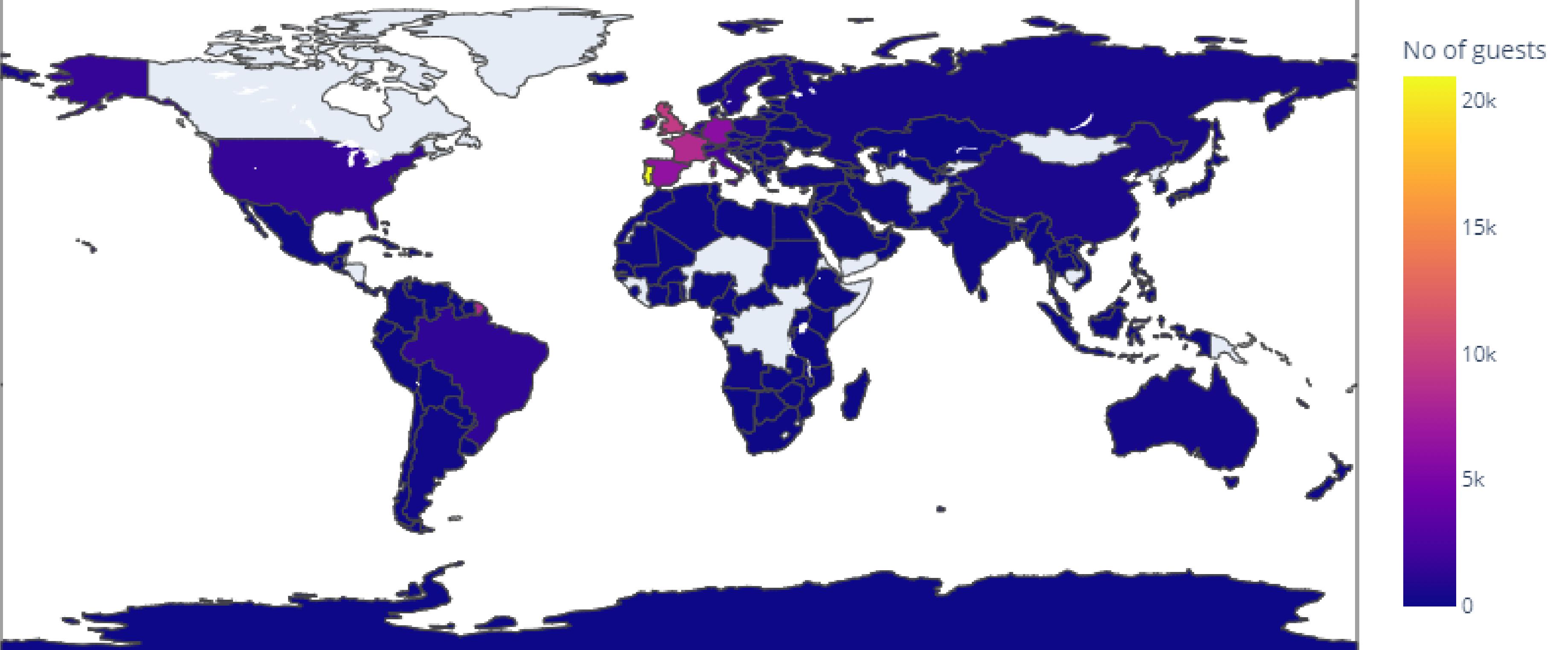
How long do people stay at the hotels?

...

Exploratory Data Analysis

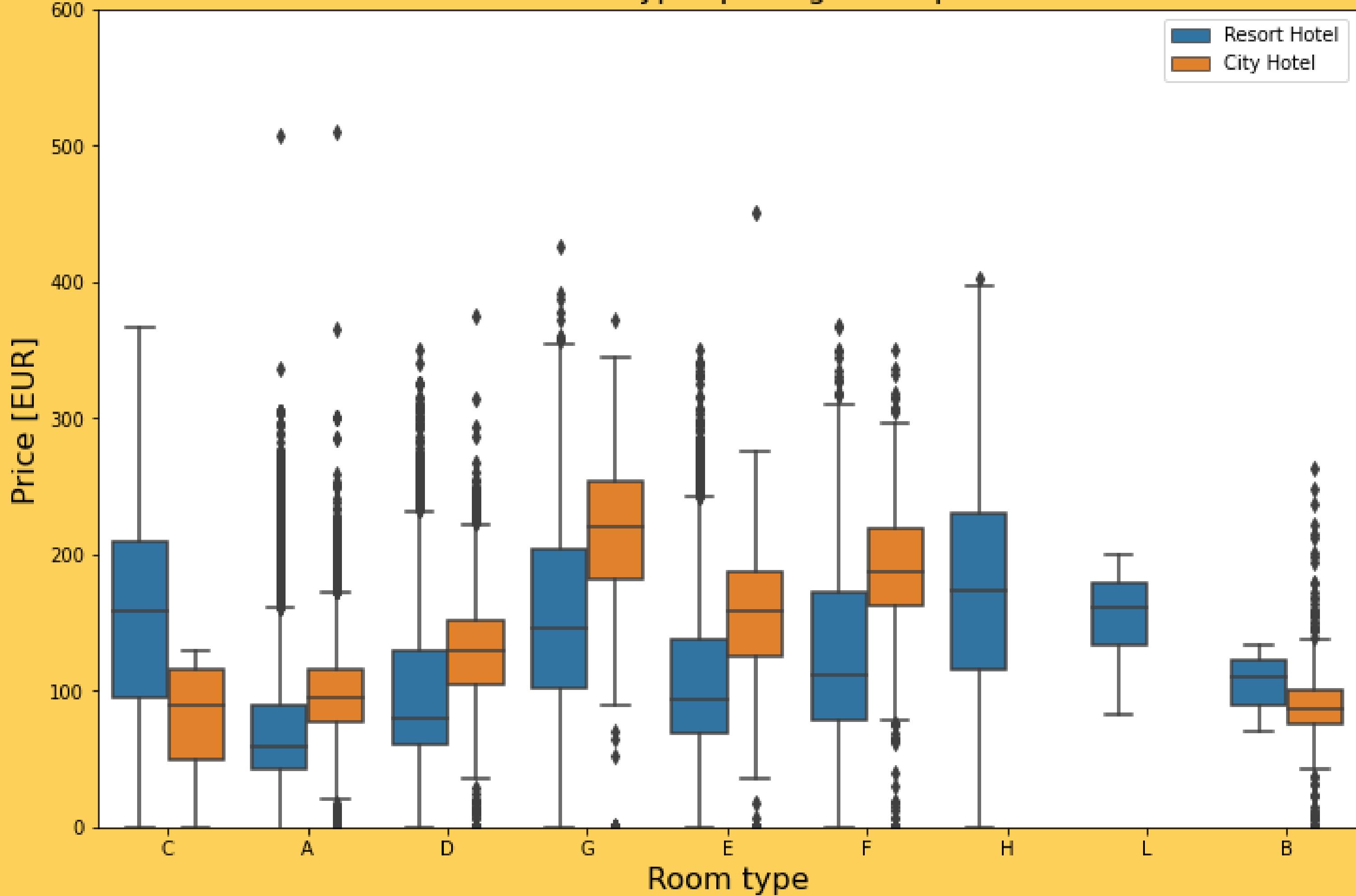
To answer business question

...

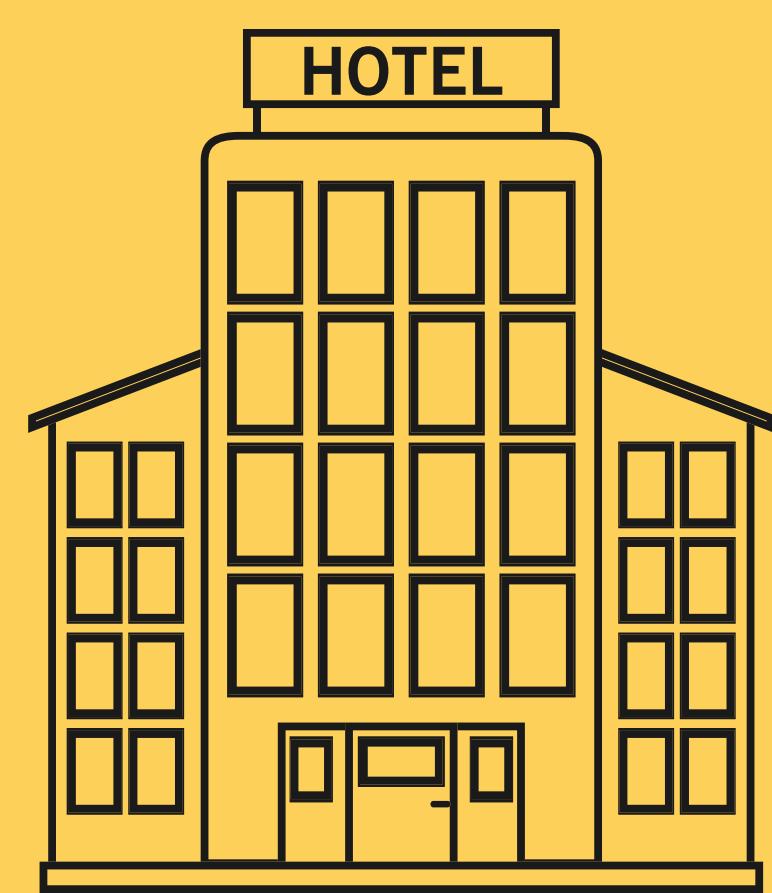


Analysing Home Country Guest with Spatial Analysis

Price of room types per night and person



**AVERAGE
PRICE FOR A
ROOM PER
NIGHT**



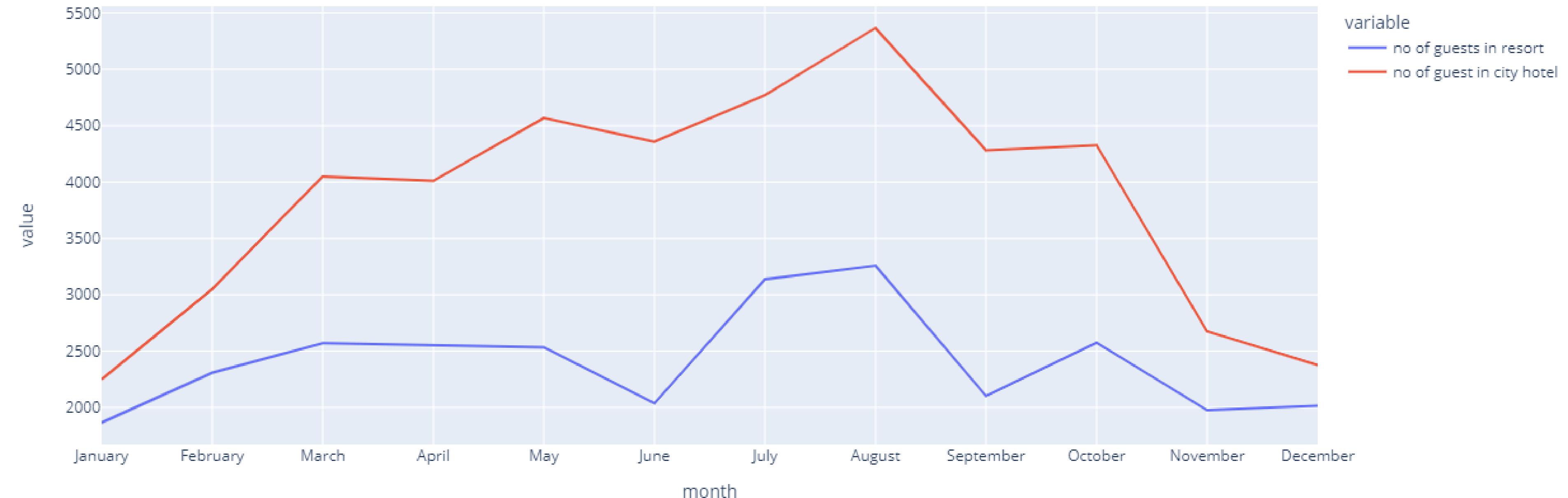
PRICE VARIATIONS PER NIGHT THROUGHOUT THE YEAR

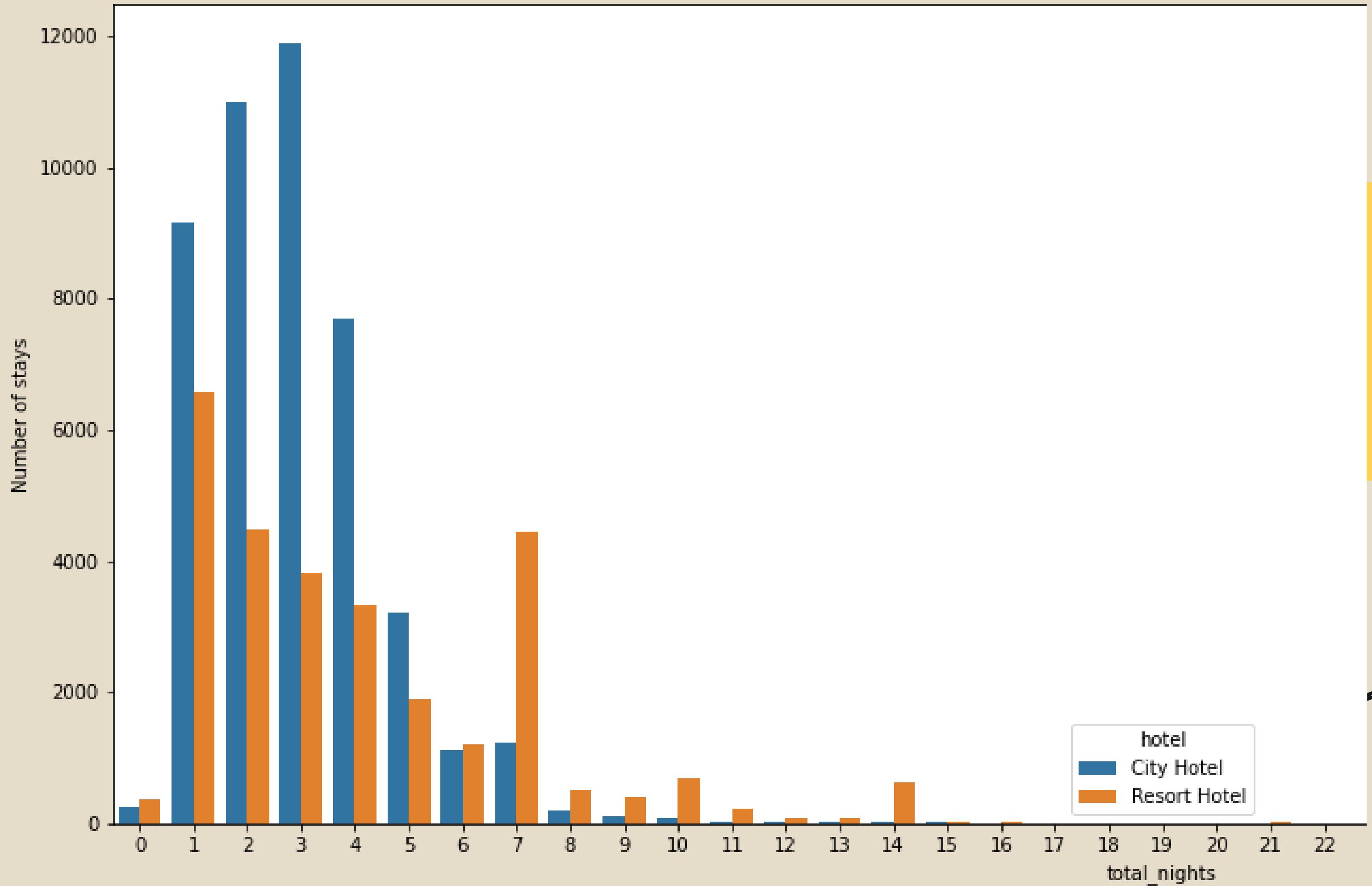
Room price per night over the Months



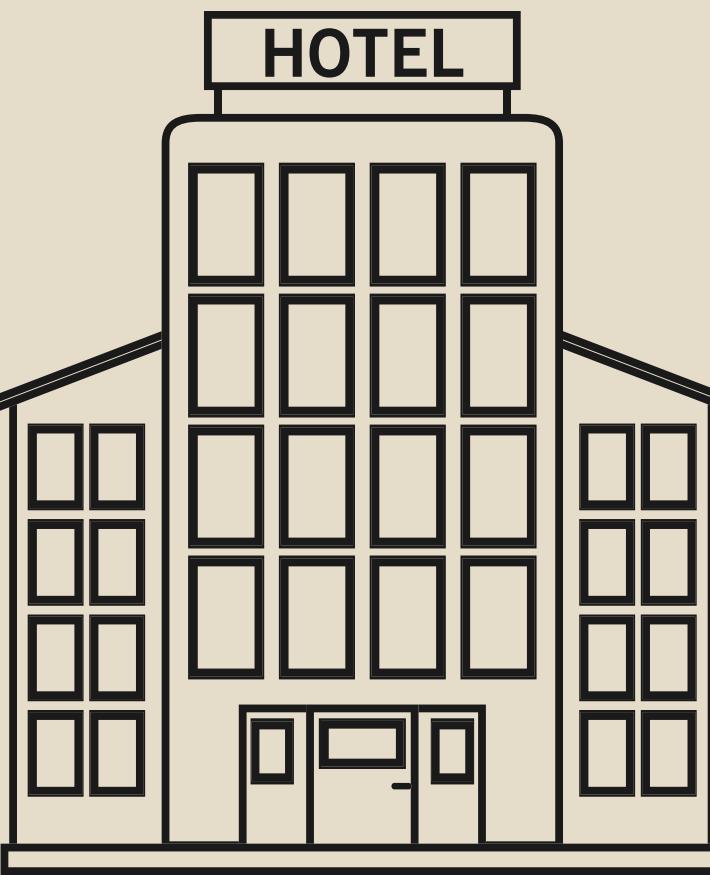
THE MOST BUSY MONTH

Total no of guests per Months

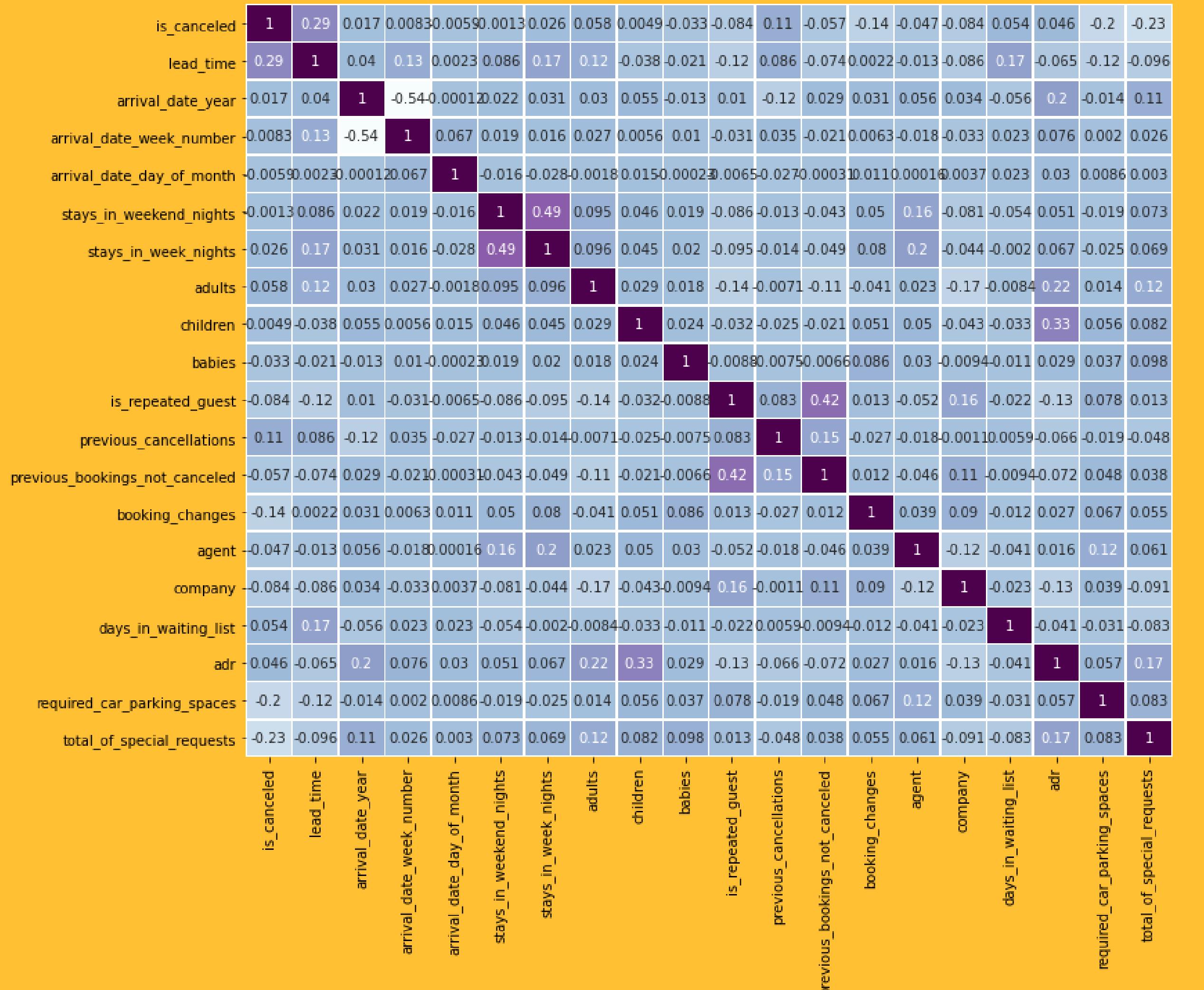




HOW LONG DO PEOPLE STAY AT THE HOTELS



SELECT IMPORTANT FEATURES USING CORELATION



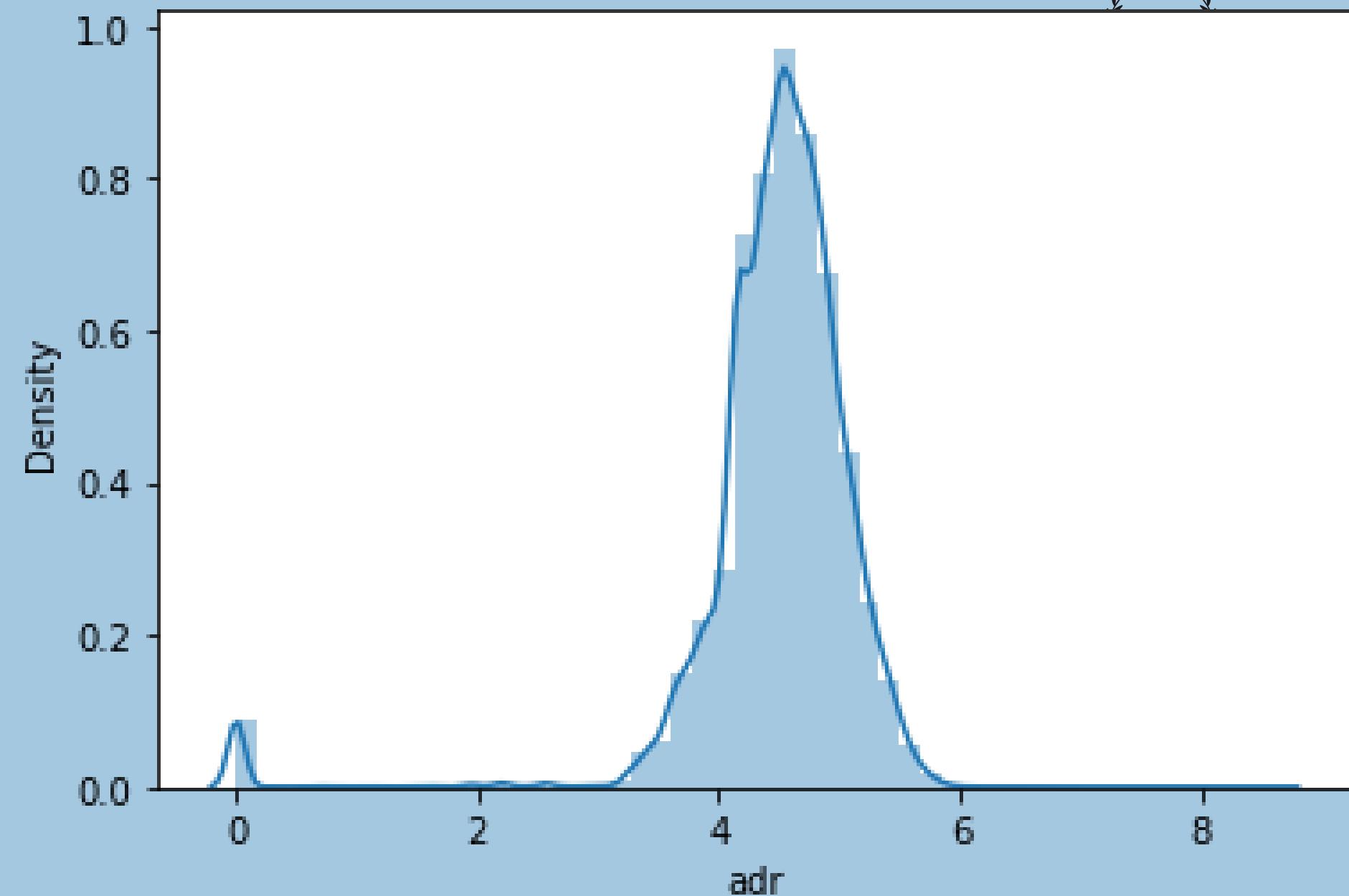
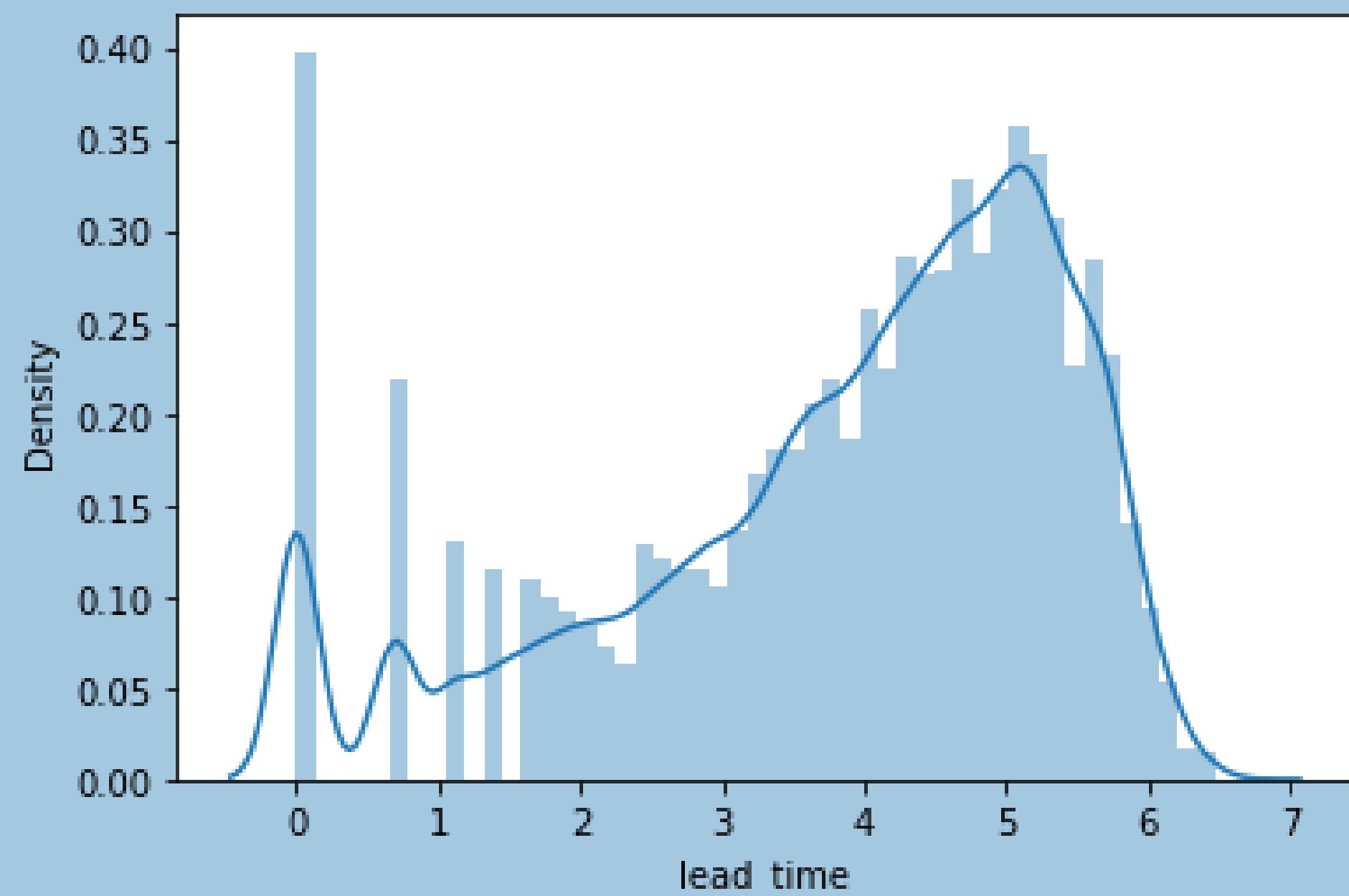
Feature Encoding



Perform Mean Encoding Technique

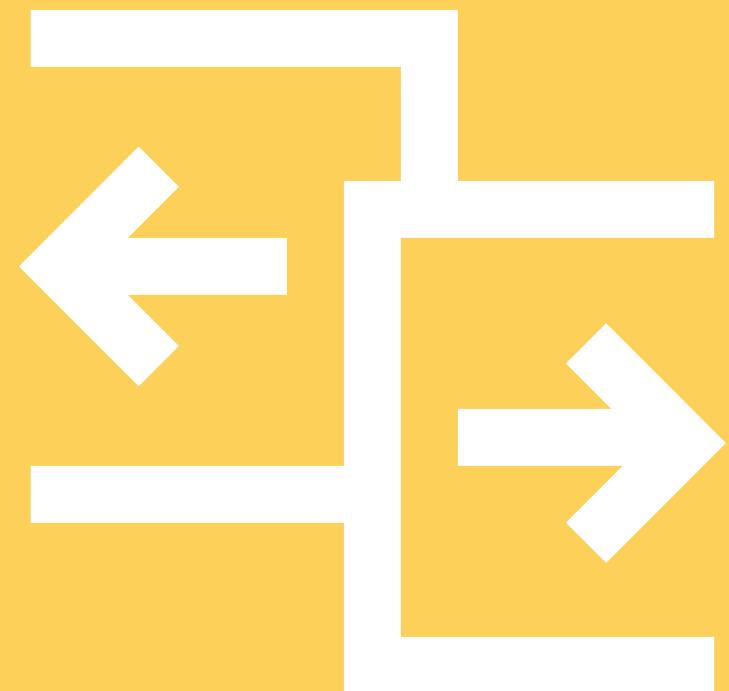
The data to be processed is contain 29 columns with 119210 rows





Handling Outlier

SEPARATE DEPENDENT & INDEPENDENT FEATURES



`~ IS_CANCELED`

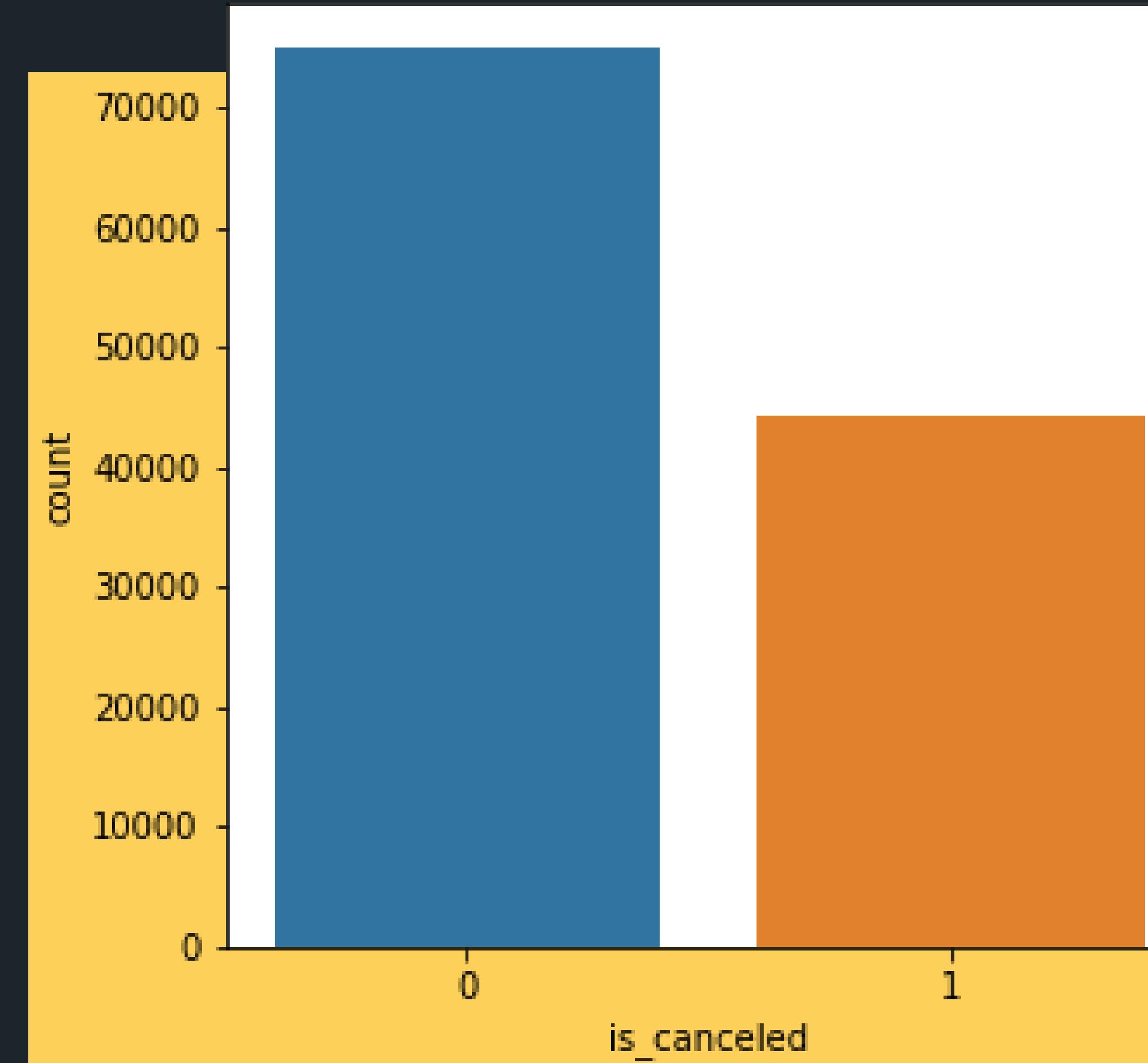
`IS_CANCELED`

IMBALANCE DATA

HANDLING WITH SMOTE METHOD

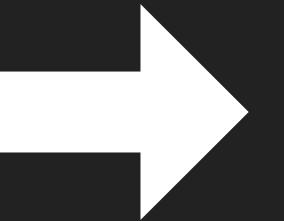
Final data after handling with SMOTE Method :

YES (75.010) and NO (75.010)



Feature Importance

```
from sklearn.linear_model import Lasso  
from sklearn.feature_selection import  
SelectFromModel
```



Total Features: 28
Selected Features: 17

...

Splitting Dataset & Build Model

...

Model	Accuracy Score	Cross Validation Score	F1 Score
Logistic Regression	0.655	0.639	0.655
Naive Bayes	0.655	0.639	0.655
RandomForest	0.968	0.767	0.968
Decision Tree	0.955	0.777	0.955
KNN	0.960	0.899	0.960



GRID SEARCH CV

Parameters (max_depth : 5, n_estimators: 300)

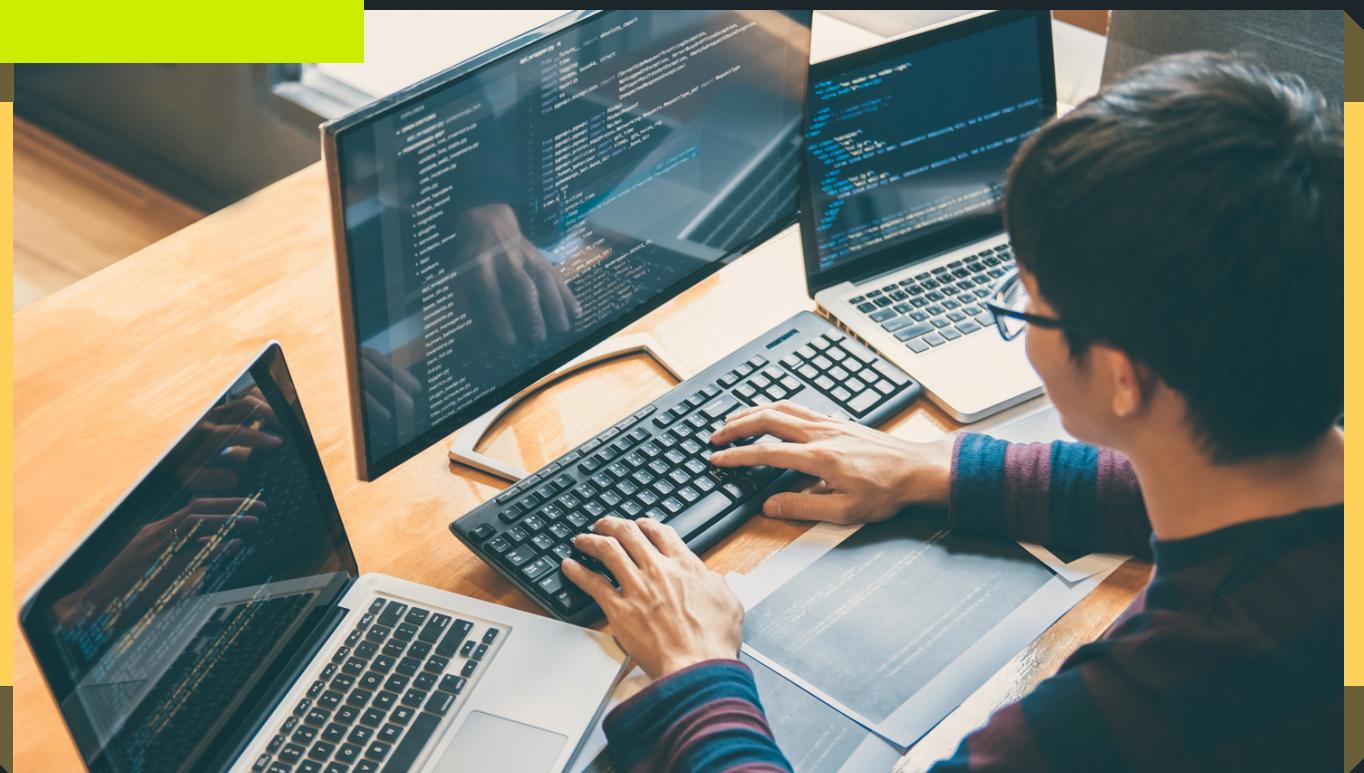
Accuracy : 0.775

HYPER TUNING PARAMETER

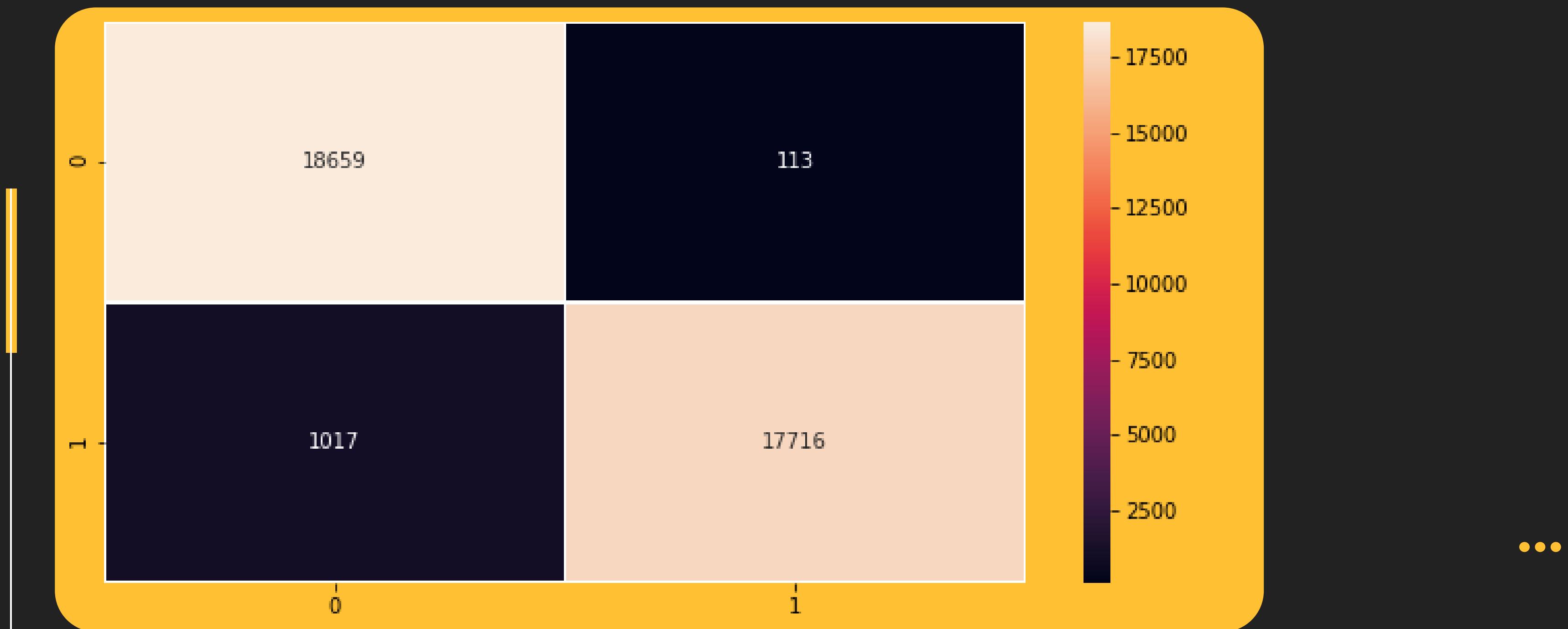
RANDOMIZED SEARCH CV

Parameters(max_depth: 5, n_estimators: 500)

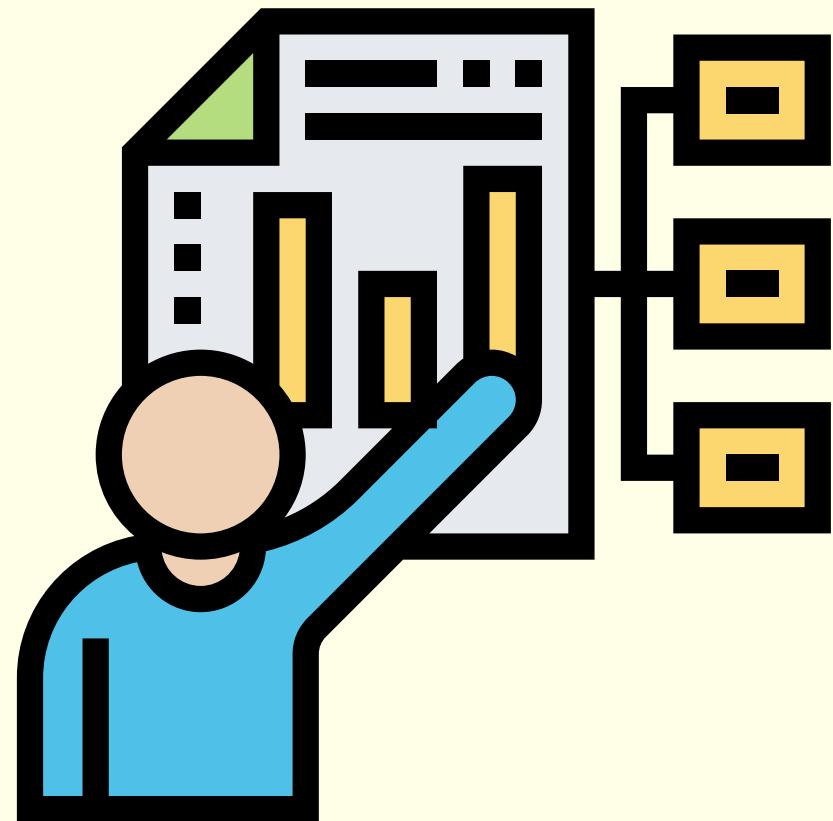
Accuracy : 0.773



Random Forest Confusion Matrix



RESULT SUMMARY



CONCLUSION

To make predictions, the best model that can be used is Random Forest. With this model, we can make predictions with the F1 Score level reaching 96.8%. This score can be achieved after passing the smote method to handle imbalance data and selecting the best 17 features.

SUGGETION

1. Focus ads on countries that visit the least or those that visit the most
2. Make adjustments to room prices from resort hotels or city hotels in order to get maximum profit
3. Holding various promotions or interesting events on a busy month.
4. Provide special prices if visitors choose to stay long enough.

Other Project



Loan Eligibility



Anime Score Prediction



Unemployment and Digital Skill
Demand Analysis



Clustering the Countries for
HELP International



COVID-19 Interactive Dashboard
with R Shiny

...

Thank You!

-END-

Linkedin: [linkedin.com/in/rikasahriana](https://www.linkedin.com/in/rikasahriana)

Email: rikasahriana28@gmail.com