



# **CANCELLATION STATUS OF HOTEL BOOKING BY USING ML ALGORITHMS**

MOGILISHETTY RAMU  
119ME0026

# AGENDA

- Problem statement
- Given data
- Visualization
- Steps we have done
- Chosen Model
- Conclusion



# PROBLEM STATEMENT:

## **Hotel Booking Cancellation prediction Problem**

Classification: Binary Hotel booking cancellation prediction is a classic problem for which many data analysis techniques must be learned and applied in order to create the best.

Classification model. Given a dataset containing information about applicants for hotel bookings as well as the status of the booking application, whether approved or not. A binary classification model must be built with maximum accuracy.



## GIVEN DATA:

Number of hotels booking  
Cancelled (Data Set) and Data  
types.

# DATA SET TABLE:

1 to 10 of 20000 entries

Filter

index	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	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_cancelled	reserved_room_type		
0	Resort Hotel	0	342	2015	July		27	1	0	0	2	0.0	0	BB	PRT	Direct	1	0	0	2	0.0	0	BB	PRT	Direct	Direct	0	0	0	C
1	Resort Hotel	0	737	2015	July		27	1	0	0	2	0.0	0	BB	PRT	Direct	1	0	0	2	0.0	0	BB	PRT	Direct	Direct	0	0	0	C
2	Resort Hotel	0	7	2015	July		27	1	0	1	1	0.0	0	BB	GBR	Direct	1	0	1	1	0.0	0	BB	GBR	Direct	Direct	0	0	0	A
3	Resort Hotel	0	13	2015	July		27	1	0	1	1	0.0	0	BB	GBR	Corporate	1	0	1	1	0.0	0	BB	GBR	Corporate	Corporate	0	0	0	A
4	Resort Hotel	0	14	2015	July		27	1	0	2	2	0.0	0	BB	GBR	Online TA	1	0	2	2	0.0	0	BB	GBR	Online TA	TA/TO	0	0	0	A
5	Resort Hotel	0	14	2015	July		27	1	0	2	2	0.0	0	BB	GBR	Online TA	1	0	2	2	0.0	0	BB	GBR	Online TA	TA/TO	0	0	0	A
6	Resort Hotel	0	0	2015	July		27	1	0	2	2	0.0	0	BB	PRT	Direct	1	0	2	2	0.0	0	BB	PRT	Direct	Direct	0	0	0	C
7	Resort Hotel	0	0	2015	July		27	1	0	2	2	0.0	0	FB	PRT	Direct	1	0	2	2	0.0	0	FB	PRT	Direct	Direct	0	0	0	C
8	Resort Hotel	1	85	2015	July		27	1	0	3	2	0.0	0	BB	PRT	Online TA	1	0	3	2	0.0	0	BB	PRT	Online TA	TA/TO	0	0	0	A
9	Resort Hotel	1	75	2015	July		27	1	0	3	2	0.0	0	HB	PRT	Offline TA	1	0	3	2	0.0	0	HB	PRT	Offline TA/TO	TA/TO	0	0	0	D

Show 10 per page

121010010001000

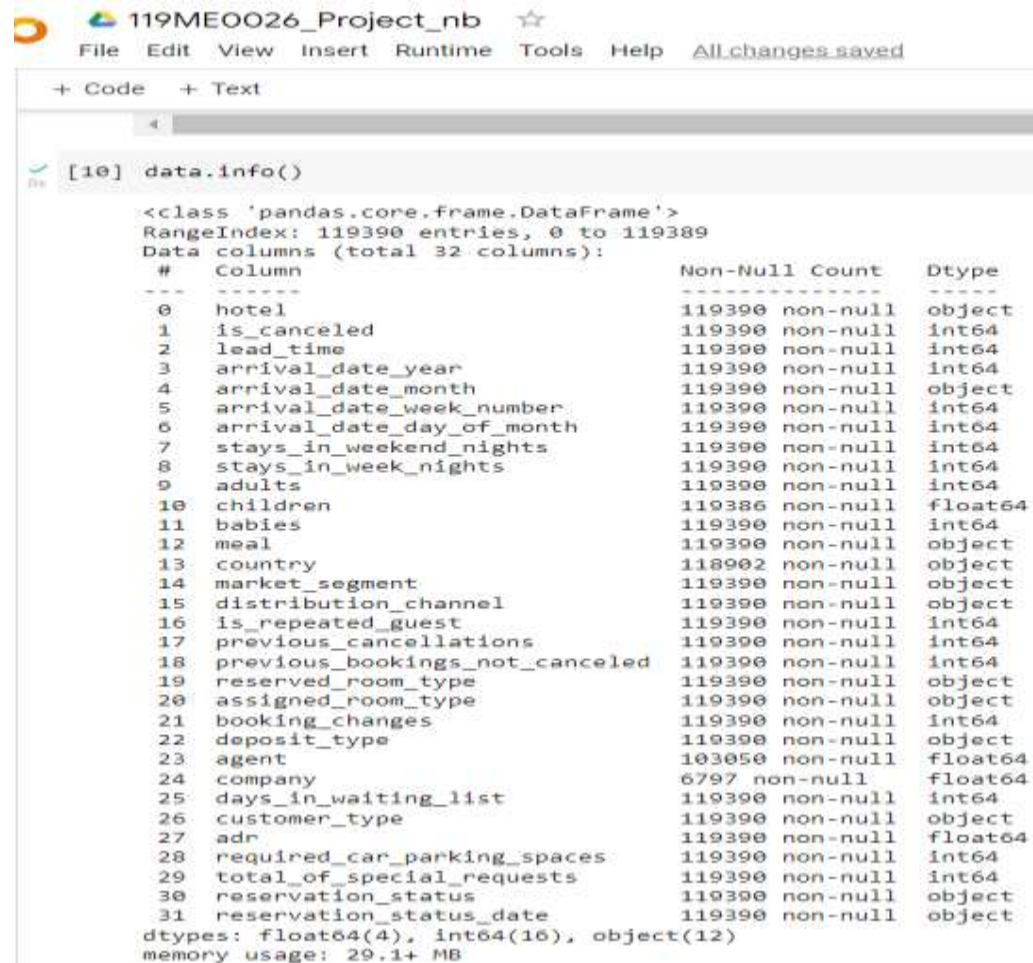
Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Warning: Total number of columns (32) exceeds max\_columns (20) limiting to first (20) columns.

Warning: total number of rows (119300) exceeds max\_rows (20000). Limiting to first (20000) rows.

Link : "[https://raw.githubusercontent.com/Premalatha-success/Datasets/main/hotel\\_bookings.csv](https://raw.githubusercontent.com/Premalatha-success/Datasets/main/hotel_bookings.csv)"

# DATA TYPES:



119ME0026\_Project\_nb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[10] data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   hotel                                119390 non-null object
1   is_canceled                         119390 non-null int64
2   lead_time                           119390 non-null int64
3   arrival_date_year                   119390 non-null int64
4   arrival_date_month                  119390 non-null object
5   arrival_date_week_number            119390 non-null int64
6   arrival_date_day_of_month           119390 non-null int64
7   stays_in_weekend_nights             119390 non-null int64
8   stays_in_week_nights                119390 non-null int64
9   adults                              119390 non-null int64
10  children                            119386 non-null float64
11  babies                             119390 non-null int64
12  meal                                119390 non-null object
13  country                             118902 non-null object
14  market_segment                      119390 non-null object
15  distribution_channel                 119390 non-null object
16  is_repeated_guest                   119390 non-null int64
17  previous_cancellations               119390 non-null int64
18  previous_bookings_not_canceled       119390 non-null int64
19  reserved_room_type                  119390 non-null object
20  assigned_room_type                  119390 non-null object
21  booking_changes                      119390 non-null int64
22  deposit_type                        119390 non-null object
23  agent                               103050 non-null float64
24  company                             6797 non-null float64
25  days_in_waiting_list                119390 non-null int64
26  customer_type                       119390 non-null object
27  adr                                  119390 non-null float64
28  required_car_parking_spaces         119390 non-null int64
29  total_of_special_requests            119390 non-null int64
30  reservation_status                  119390 non-null object
31  reservation_status_date              119390 non-null object
dtypes: float64(4), int64(10), object(12)
memory usage: 29.1+ MB
```



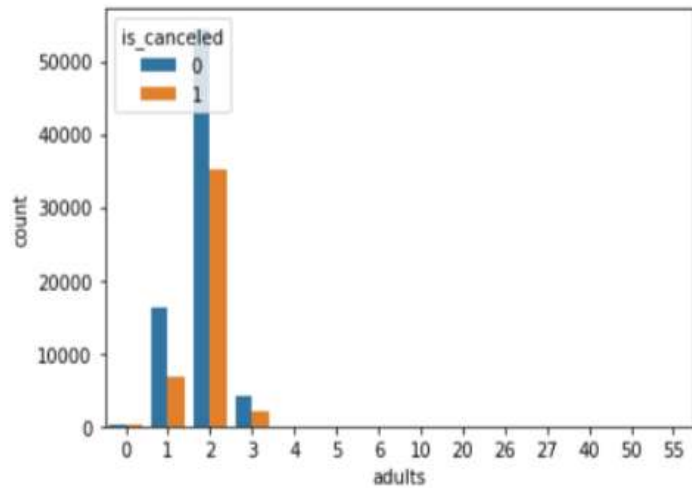
# VISUALIZATION WITH INFERENCE



# VISUALIZING FEATURE TO FEATURE:

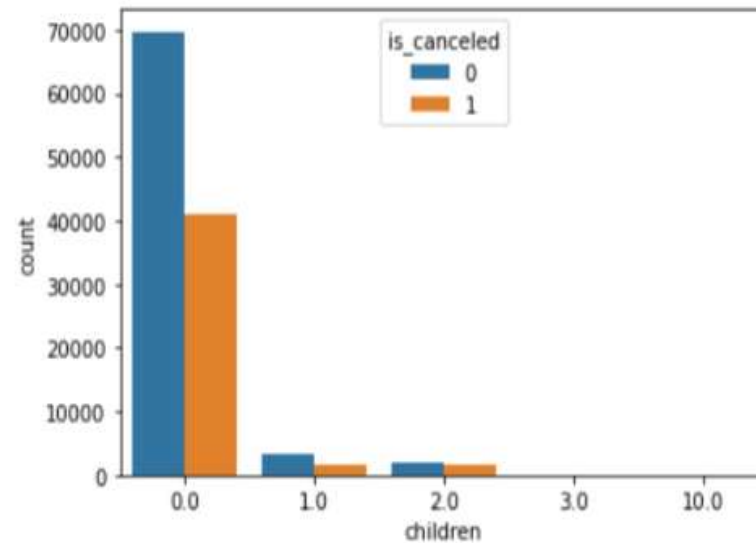
```
sns.countplot(x="adults",hue="is_canceled",data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f120dd96f10>
```



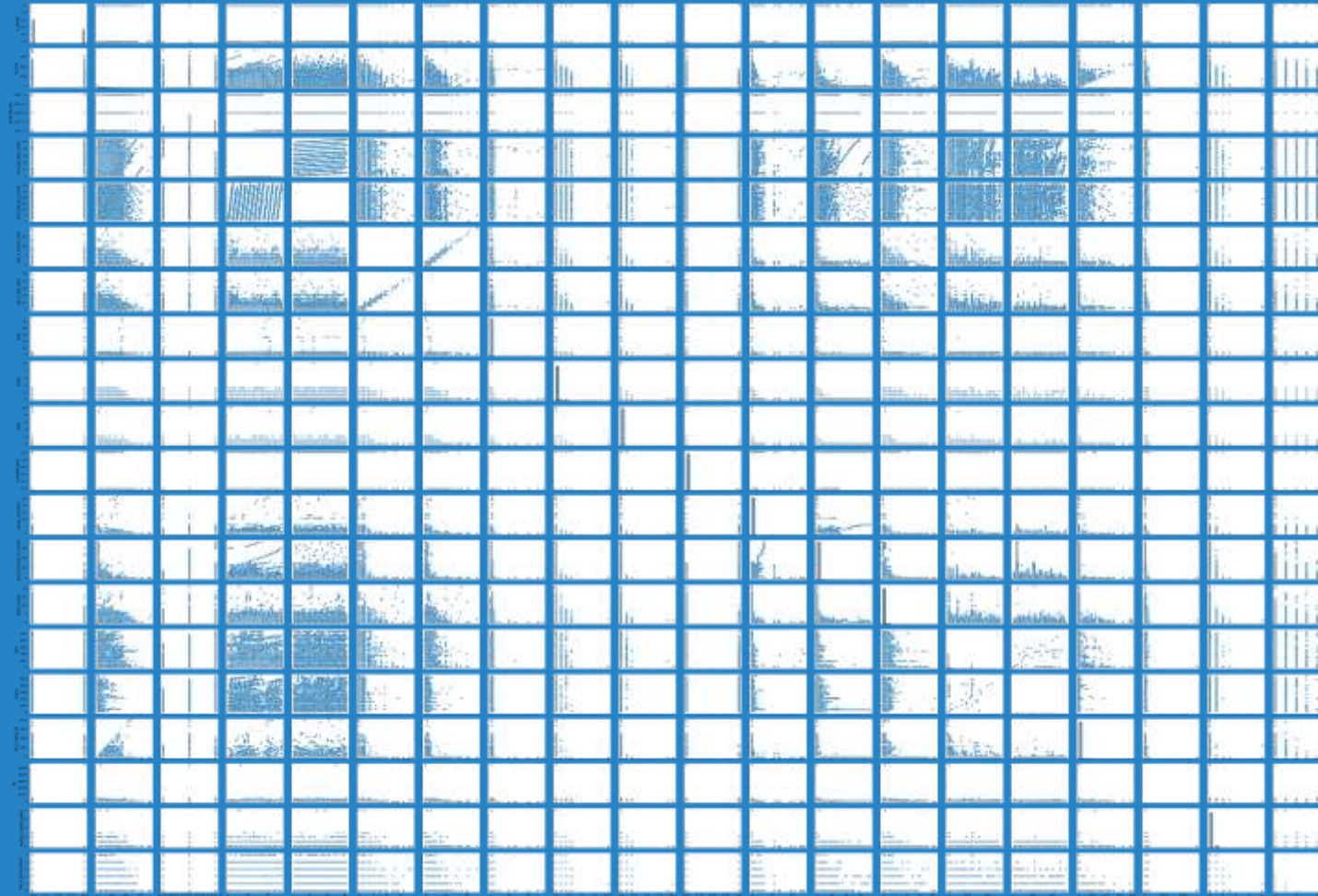
```
sns.countplot(x="children",hue="is_canceled",data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f120dc73890>
```

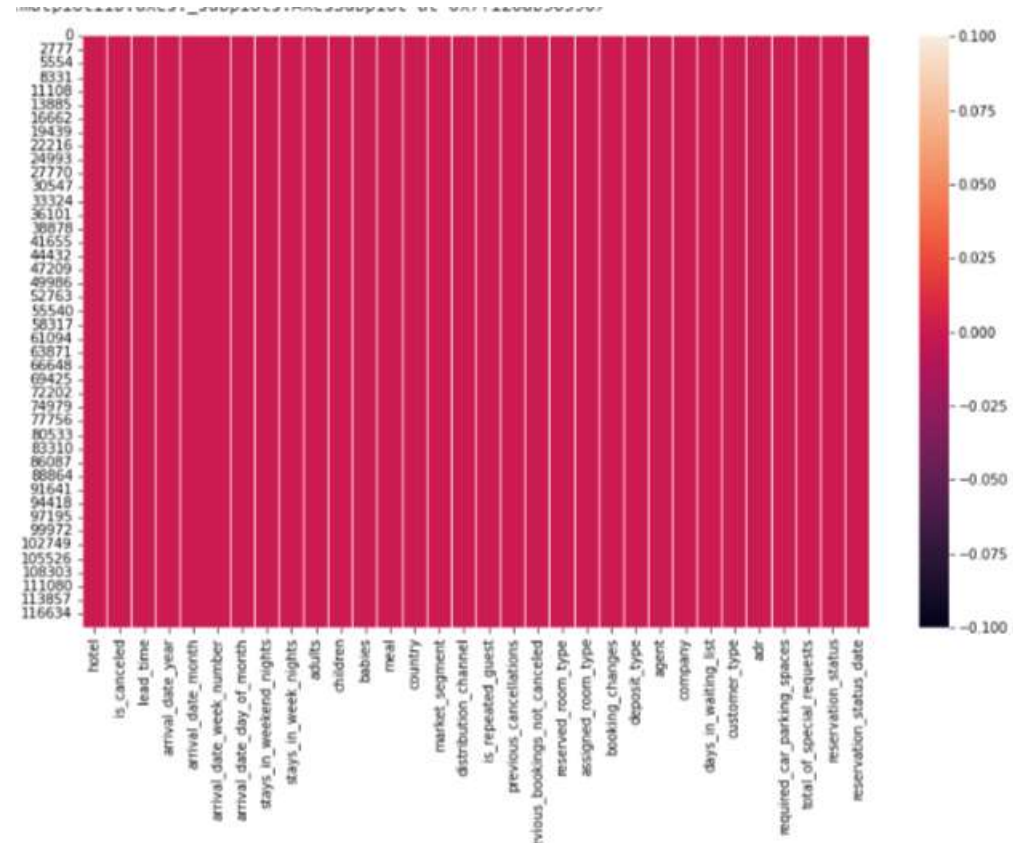
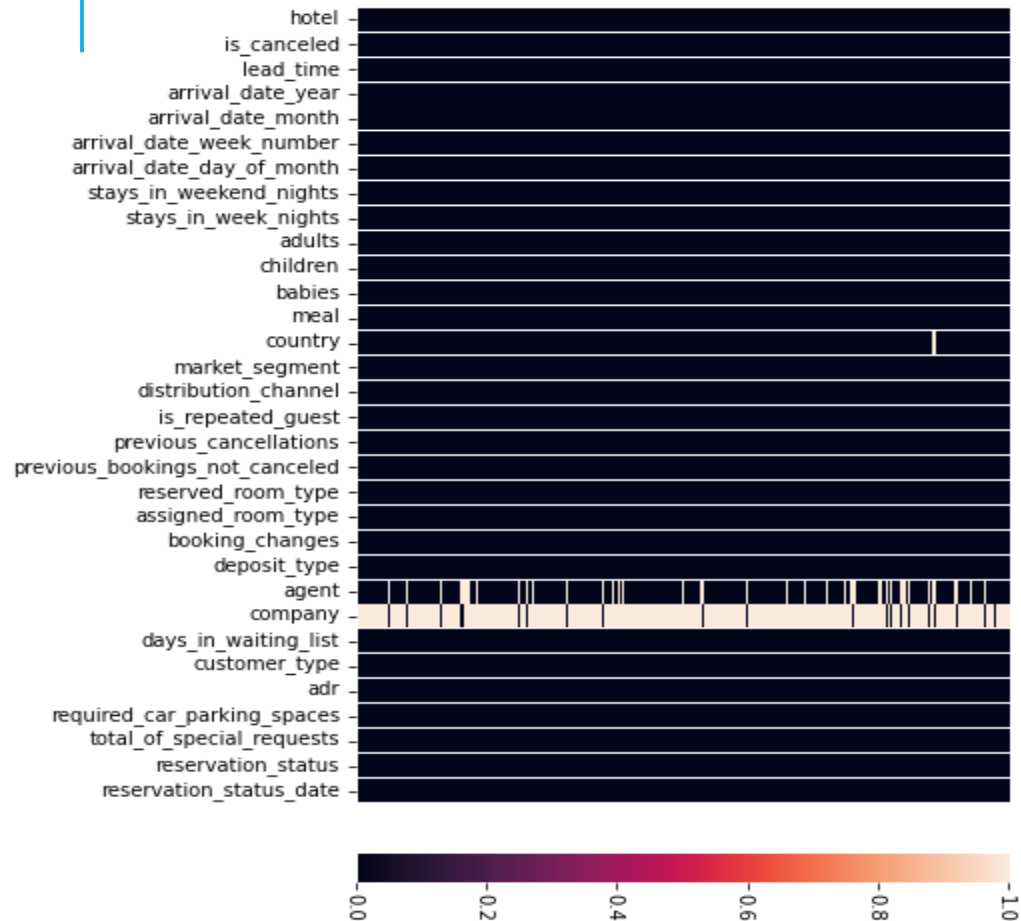




# THE PAIR PLOT:



# VISUALIZING NULL VALUES:



- Replaced Null values with median for int and float and mode for object type.



# STEPS WE HAVE DONE:



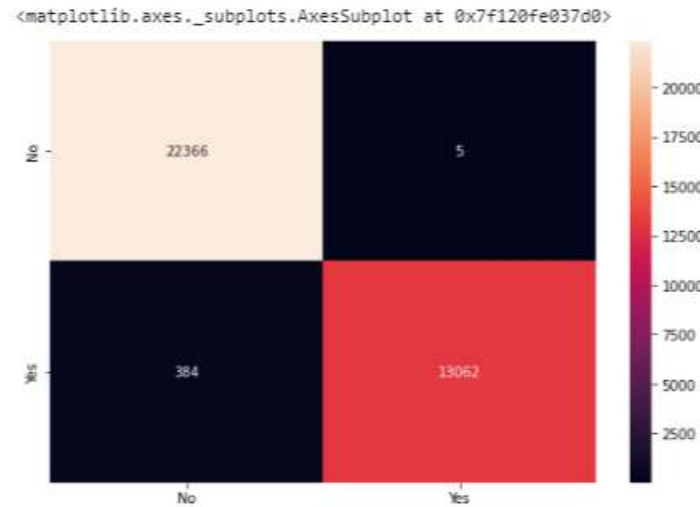
# CHOSEN MODEL: **DECISION TREE**

## THE REASON WHY WE CHOOSE DECISION TREE IS

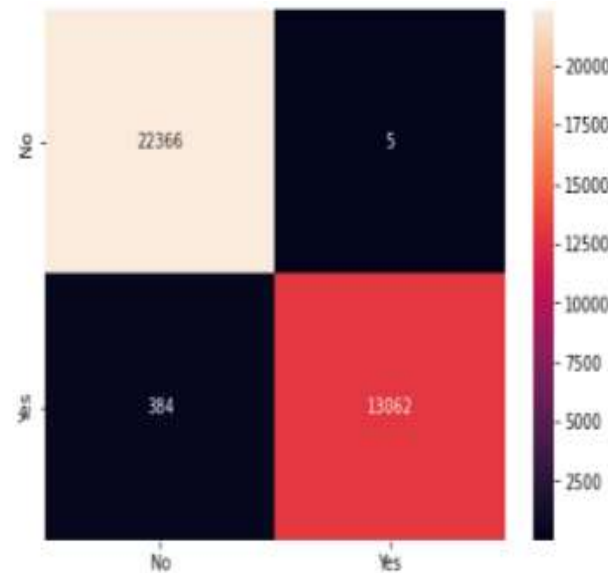
- In this experiment, I used three Machine Learning Algorithms: Logistic Regression, K-Nearest Neighbors, and Decision Tree Algorithms.
- But, in the end, I chose the Decision tree model because its accuracy rate is 100%.

# CONFUSION MATRICES OF ALGORITHMS:

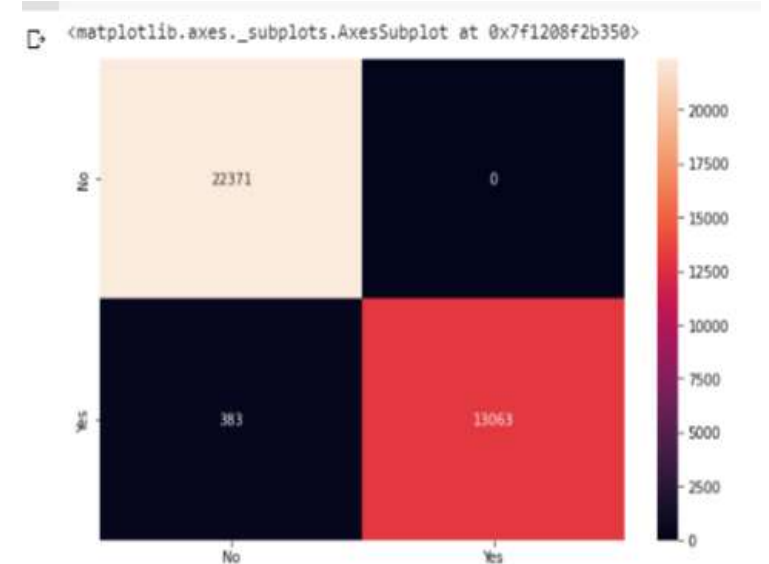
## Logistic Regression



## K-Nearest Neighbors



## Decision Tree



- In this project we used three Machine learning classification algorithms to know the cancellation status of hotel rooms by using given data set.
- After the visualization of features in the data set by using training data and testing data we come to know among the three algorithms the accuracy rate & precision is different they are –

1. Logistic regression:

Training Data: 0.98996769915176

Testing data:0.9891392355585337

2. KNN:

Training Data:0.9193639093965754

Testing data:0.88061535025226733

3. Decision Tree:

Training Data: 1.0

Testing Data: 1.0

## CONCLUSION:

- So, we come to know that Decision Tree model classifier is working more accurate in booking of services from Hotel.



THANK YOU

M. RAMU  
119ME0026.