Hotel reservation Presentation

By- T.Bhanu Teja

119ME0003

Project objective

Hotel Reservation: Train the machine learning model to predict, if a customer would cancel a hotel reservation given a dataset containing information on reservations made by customers in various hotels.

Attributes in Data set

In formulating the Hotel reservation, the factors are :

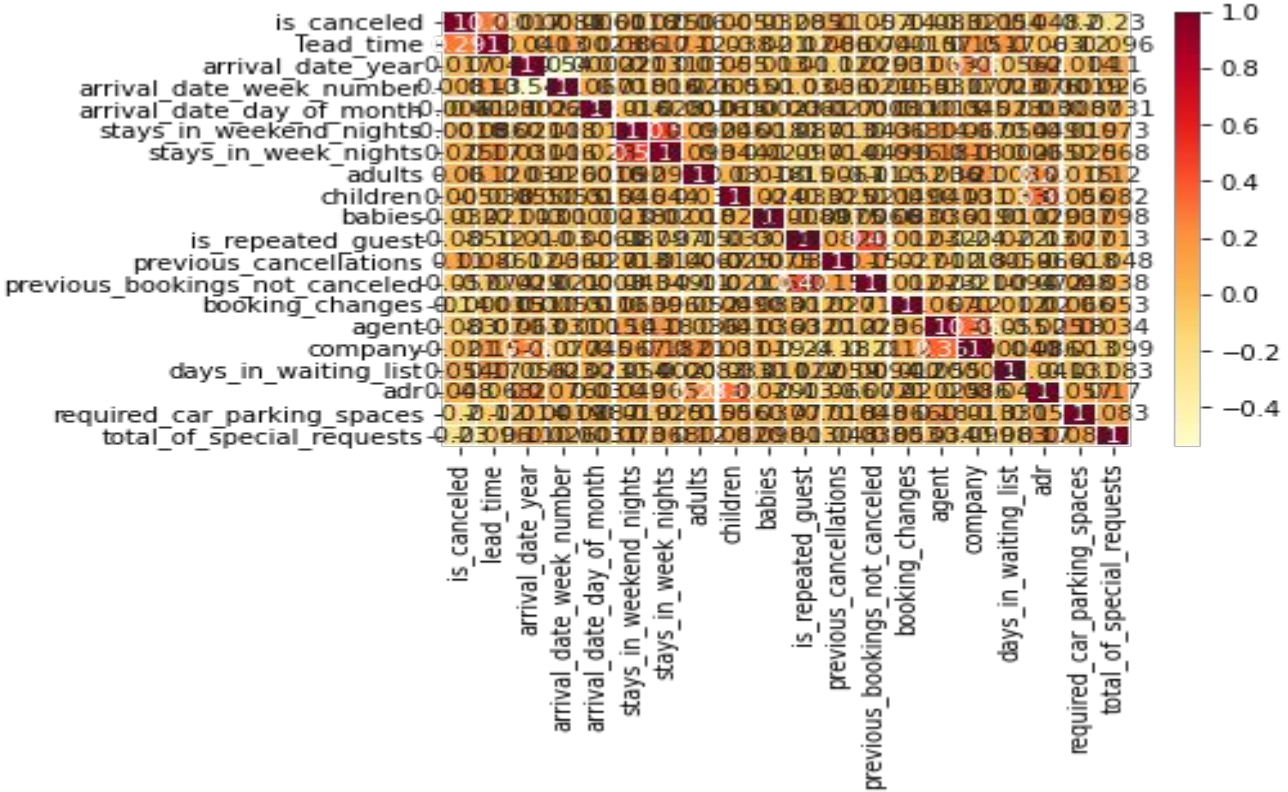
Arrival date , Stays in week nights ,market segment,distribution channel, repeated guest, previous cancellations, previous bookings not canceled,reserved rooms, assigned rooms, deposits ,adults,children, agent, company,Days in waiting list, customer type, required car parking spaces, total of special requests, reservation status. Dataset contains 119390 rows and 32 columns which describes the all booking details.

Persona 01

Data visualization

This data set contains booking information for a 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. All personally identifying information has from the data.

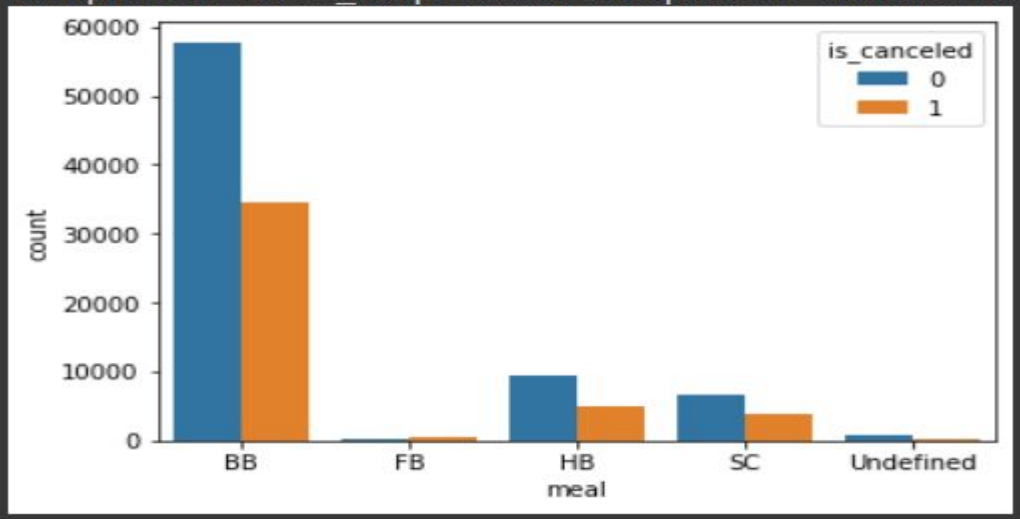
Correlation visualization



Visualizing features

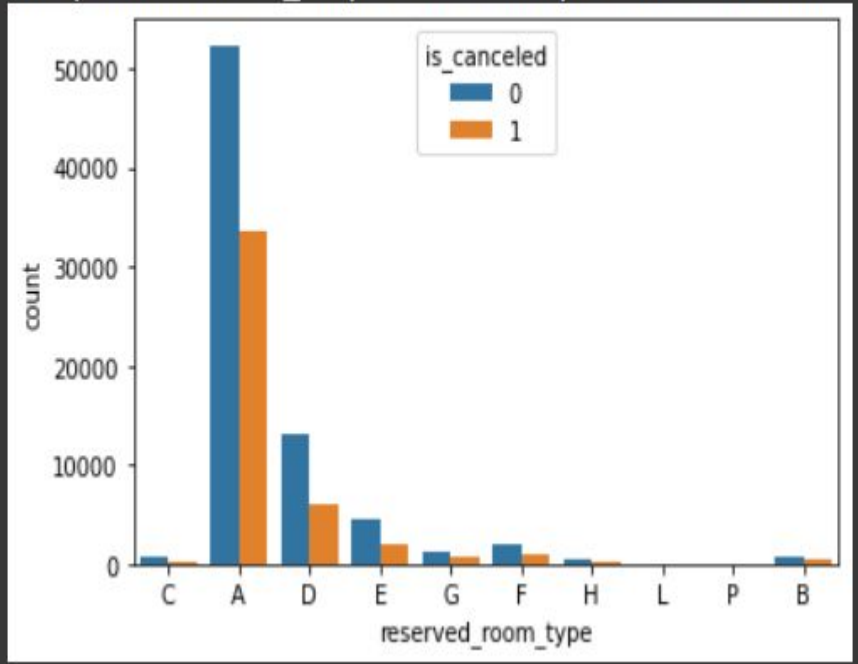
```
1 sns.countplot(x="meal",hue="is_canceled",data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f986f9ac410>
```

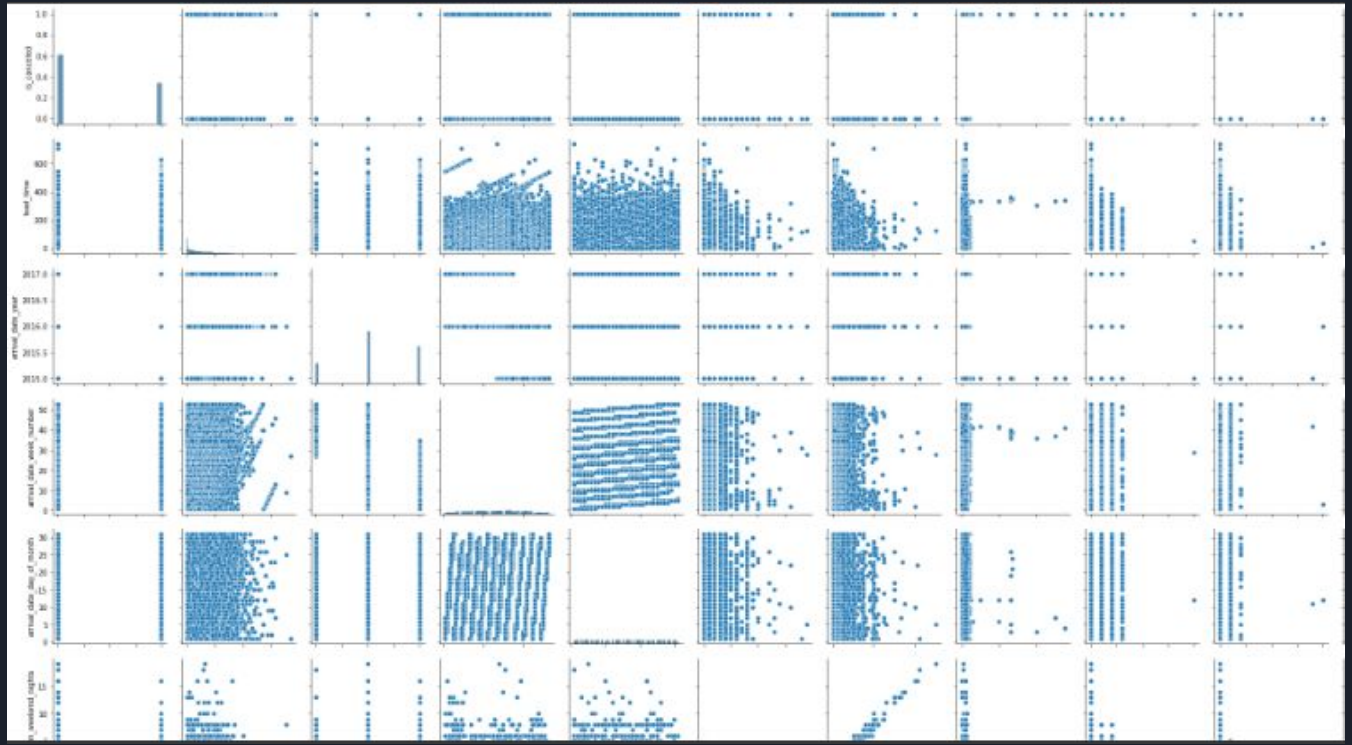


```
1 sns.countplot(x="reserved_room_type",hue="is_canceled",data=data)
```

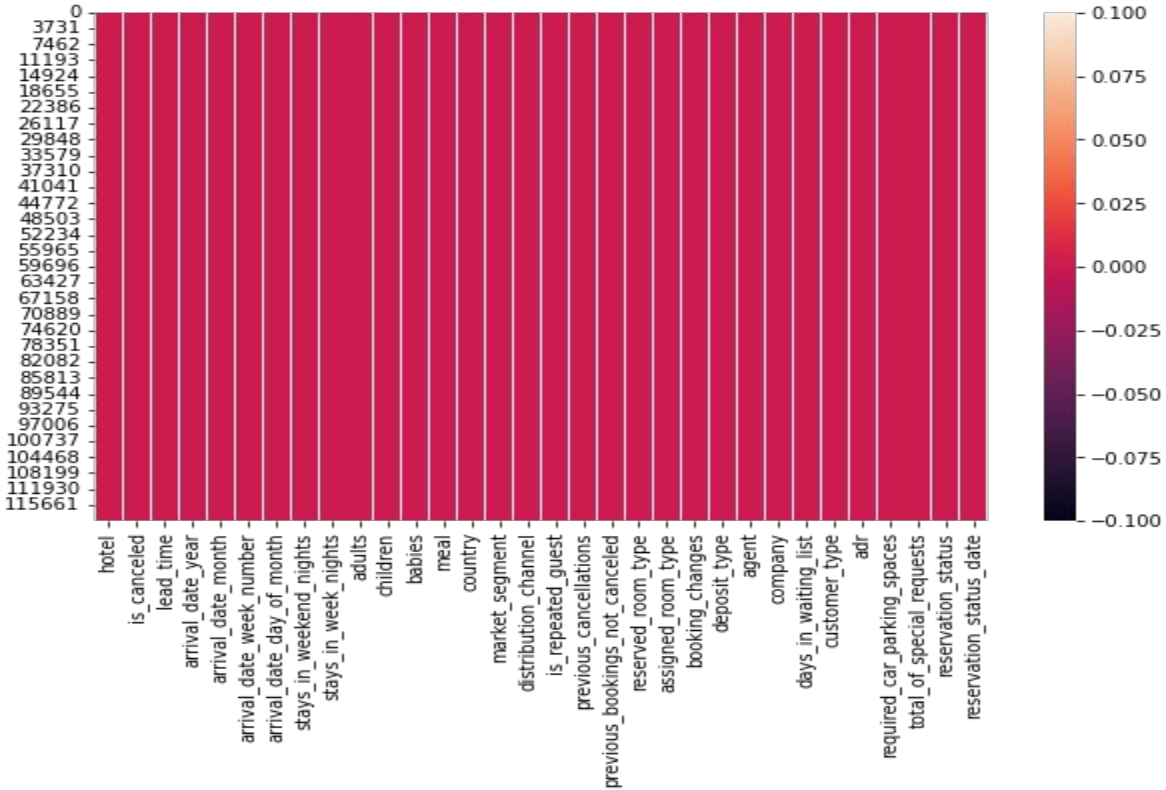
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9874b91e50>
```



Visualizing pairplot



Visualizing Null values



Steps we done

Imported packages:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import AdaBoostClassifier ,
GradientBoostingClassifier,RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report , accuracy_score ,
confusion_matrix
from sklearn.metrics import f1_score
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
import warnings
warnings.filterwarnings("ignore")
```

Now import and display the dataset

```
[69]  1 data = pd.read_csv("https://raw.githubusercontent.com/Premalatha-success/Datasets/main/hotel_bookings.csv")
      2 data.head(5)
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	

5 rows × 32 columns

Converting Data Type

```
1 data.dtypes
```

```
hotel                           object
is_canceled                      int64
lead_time                        int64
arrival_date_year                int64
arrival_date_month              object
arrival_date_week_number         int64
arrival_date_day_of_month        int64
stays_in_weekend_nights          int64
stays_in_week_nights             int64
adults                           int64
children                       float64
babies                           int64
meal                            object
country                         object
market_segment                  object
distribution_channel            object
is_repeated_guest                int64
previous_cancellations           int64
previous_bookings_not_canceled   int64
reserved_room_type              object
assigned_room_type              object
booking_changes                  int64
deposit_type                    object
agent                          float64
company                        float64
days_in_waiting_list             int64
customer_type                   object
adr                            float64
required_car_parking_spaces      int64
total_of_special_requests        int64
reservation_status              object
reservation_status_date         object
dtype: object
```

Encoding the Categorical variables

1. Hotel
2. arrival_date_month
3. meal
4. country
5. market_segment
6. distribution_channel
7. reserved_room_type
8. assigned_room_type
9. deposit_type
10. customer_type
11. reservation_status
12. reservation_status_date

Steps we done

Step 3: Explore the data-shape,visualization

Step 4: X,y-->train data test data > Fit the model with training data predict with the test data

Step-5: Applied all the algorithms to predict best classifier

Step-6:Visualized all the confusion metrics.

Accuracy using algorithms

Logistic Regression:

Training Accuracy : 0.9899967692915176

Testing Accuracy : 0.9891392355585337

Decision Tree:

Training Accuracy : 1.0

Testing Accuracy : 1.0

Accuracy using algorithms

Bagging Classifier:

Training Accuracy : 1.0

Testing Accuracy : 1.0

Boosting Classifier:

Training Accuracy : 1.0

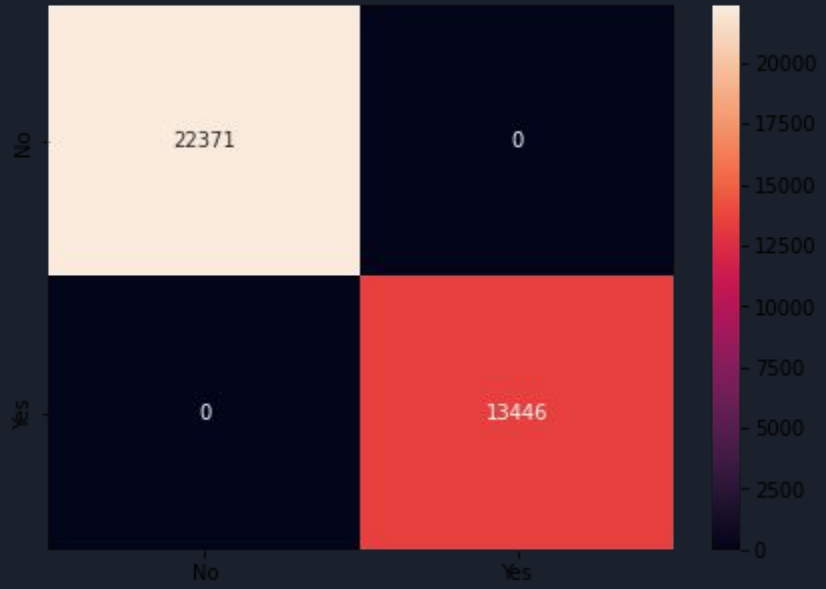
Testing Accuracy : 1.0

Chosen Model: Decision Tree

I experimented with two machine learning algorithms: logistic regression and decision trees.

-But in the end, I decided on the Decision Tree Model because it has a 100% accuracy rate.

Confusion Matrices of Algorithms of Decision tree:



Conclusion

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 –

Decision Tree:

Training Data: 1.0

Testing Data: 1.0

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

Thank you