hotel booking project-Copy1

May 25, 2021

Hotel booking project is a project that will predict how likely it is for a customer to cancel their hotel booking

1 data cleaning

[54]:

```
#importing libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[55]:

```
#read the data

hotel = pd.read_csv("C:/Users/hp/Downloads/hotel_bookings.csv")
```

[56]:

```
#first 5 rows

hotel.head()
```

[56]:

```
          hotel  is_canceled  lead_time  arrival_date_year arrival_date_month  \
0  Resort Hotel            0        342               2015               July
1  Resort Hotel            0        737               2015               July
2  Resort Hotel            0          7               2015               July
3  Resort Hotel            0         13               2015               July
4  Resort Hotel            0         14               2015               July

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1
1                        27                          1
2                        27                          1
3                        27                          1
4                        27                          1

   stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
0                        0                     0       2       0.0       0
1                        0                     0       2       0.0       0
```

```
2                        0                     1       1       0.0       0
3                        0                     1       1       0.0       0
4                        0                     2       2       0.0       0

  meal country market_segment distribution_channel  is_repeated_guest  \
0   BB     PRT         Direct               Direct                  0
1   BB     PRT         Direct               Direct                  0
2   BB     GBR         Direct               Direct                  0
3   BB     GBR      Corporate            Corporate                  0
4   BB     GBR      Online TA                TA/TO                  0

   previous_cancellations  previous_bookings_not_canceled reserved_room_type  \
0                       0                               0                  C
1                       0                               0                  C
2                       0                               0                  A
3                       0                               0                  A
4                       0                               0                  A

  assigned_room_type  booking_changes deposit_type  agent  company  \
0                  C                3   No Deposit    NaN      NaN
1                  C                4   No Deposit    NaN      NaN
2                  C                0   No Deposit    NaN      NaN
3                  A                0   No Deposit  304.0      NaN
4                  A                0   No Deposit  240.0      NaN

   days_in_waiting_list customer_type   adr  required_car_parking_spaces  \
0                     0     Transient   0.0                            0
1                     0     Transient   0.0                            0
2                     0     Transient  75.0                            0
3                     0     Transient  75.0                            0
4                     0     Transient  98.0                            0

   total_of_special_requests reservation_status reservation_status_date
0                          0          Check-Out                7/1/2015
1                          0          Check-Out                7/1/2015
2                          0          Check-Out                7/2/2015
3                          0          Check-Out                7/2/2015
4                          1          Check-Out                7/3/2015
```

[57]:

```
# number os rows and number of columns

hotel.shape
```

[57]:

```
(119390, 32)
```

[58]:

```
#checking for missing data
```

```
hotel.isnull().sum()
```

[58]:

```
hotel                                  0
is_canceled                            0
lead_time                              0
arrival_date_year                      0
arrival_date_month                     0
arrival_date_week_number               0
arrival_date_day_of_month              0
stays_in_weekend_nights                0
stays_in_week_nights                   0
adults                                 0
children                               4
babies                                 0
meal                                   0
country                              488
market_segment                         0
distribution_channel                   0
is_repeated_guest                      0
previous_cancellations                 0
previous_bookings_not_canceled         0
reserved_room_type                     0
assigned_room_type                     0
booking_changes                        0
deposit_type                           0
agent                              16340
company                           112593
days_in_waiting_list                   0
customer_type                          0
adr                                    0
required_car_parking_spaces            0
total_of_special_requests              0
reservation_status                     0
reservation_status_date                0
dtype: int64
```

[59]:

```
#dealing with missing values

def data_clean(hotel):
    hotel.fillna(0,inplace=True)
    print(hotel.isnull().sum())
```

[60]:

```
#calling the function and now we dont have any null values

data_clean(hotel)
```

```
hotel                             0
is_canceled                       0
```

```
lead_time                         0
arrival_date_year                 0
arrival_date_month                0
arrival_date_week_number          0
arrival_date_day_of_month         0
stays_in_weekend_nights           0
stays_in_week_nights              0
adults                            0
children                          0
babies                            0
meal                              0
country                           0
market_segment                    0
distribution_channel              0
is_repeated_guest                 0
previous_cancellations            0
previous_bookings_not_canceled    0
reserved_room_type                0
assigned_room_type                0
booking_changes                   0
deposit_type                      0
agent                             0
company                           0
days_in_waiting_list              0
customer_type                     0
adr                               0
required_car_parking_spaces       0
total_of_special_requests         0
reservation_status                0
reservation_status_date           0
dtype: int64
```

[61]:

```
#display the columns

hotel.columns
```

[61]:

```
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', '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', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

[62]:

```
#finding the unique instances in the three categories of people

list =['adults', 'children', 'babies']
for val in list:
    print('{} has uniques values as {}'.format(val,hotel[val].unique()))
```

```
adults has uniques values as [ 2  1  3  4 40 26 50 27 55  0 20  6  5 10]
children has uniques values as [ 0.  1.  2. 10.  3.]
babies has uniques values as [ 0  1  2 10  9]
```

[63]:

```
#creating a filter for the zeros that exist in all the three categories

filter =(hotel["adults"]== 0) & (hotel["children"]==0) & (hotel['babies']==0)
hotel[filter]
```

[63]:

```
               hotel  is_canceled  lead_time  arrival_date_year  \
2224    Resort Hotel            0          1               2015
2409    Resort Hotel            0          0               2015
3181    Resort Hotel            0         36               2015
3684    Resort Hotel            0        165               2015
3708    Resort Hotel            0        165               2015
…                …           …        …                …
115029    City Hotel            0        107               2017
115091    City Hotel            0          1               2017
116251    City Hotel            0         44               2017
116534    City Hotel            0          2               2017
117087    City Hotel            0        170               2017

       arrival_date_month  arrival_date_week_number  \
2224              October                        41
2409              October                        42
3181             November                        47
3684             December                        53
3708             December                        53
…                    …                        …
115029               June                        26
115091               June                        26
116251               July                        28
116534               July                        28
117087               July                        30

        arrival_date_day_of_month  stays_in_weekend_nights  \
2224                            6                        0
2409                           12                        0
3181                           20                        1
3684                           30                        1
```

```
3708                           30                        2
…                           …                      …
115029                         27                        0
115091                         30                        0
116251                         15                        1
116534                         15                        2
117087                         27                        0

        stays_in_week_nights  adults  children  babies meal country  \
2224                       3       0       0.0       0   SC     PRT
2409                       0       0       0.0       0   SC     PRT
3181                       2       0       0.0       0   SC     ESP
3684                       4       0       0.0       0   SC     PRT
3708                       4       0       0.0       0   SC     PRT
…                      …     …       …     …  …     …
115029                     3       0       0.0       0   BB     CHE
115091                     1       0       0.0       0   SC     PRT
116251                     1       0       0.0       0   SC     SWE
116534                     5       0       0.0       0   SC     RUS
117087                     2       0       0.0       0   BB     BRA

       market_segment distribution_channel  is_repeated_guest  \
2224        Corporate            Corporate                  0
2409        Corporate            Corporate                  0
3181           Groups                TA/TO                  0
3684           Groups                TA/TO                  0
3708           Groups                TA/TO                  0
…               …                  …                …
115029      Online TA                TA/TO                  0
115091  Complementary               Direct                  0
116251      Online TA                TA/TO                  0
116534      Online TA                TA/TO                  0
117087  Offline TA/TO                TA/TO                  0

        previous_cancellations  previous_bookings_not_canceled  \
2224                         0                               0
2409                         0                               0
3181                         0                               0
3684                         0                               0
3708                         0                               0
…                        …                             …
115029                       0                               0
115091                       0                               0
116251                       0                               0
116534                       0                               0
117087                       0                               0
```

```
       reserved_room_type assigned_room_type  booking_changes deposit_type  \
2224                    A                  I                1   No Deposit   
2409                    A                  I                0   No Deposit   
3181                    A                  C                0   No Deposit   
3684                    A                  A                1   No Deposit   
3708                    A                  C                1   No Deposit   
…                   …                …              …          …
115029                  A                  A                1   No Deposit   
115091                  E                  K                0   No Deposit   
116251                  A                  K                2   No Deposit   
116534                  A                  K                1   No Deposit   
117087                  A                  A                0   No Deposit   

        agent  company  days_in_waiting_list    customer_type     adr  \
2224      0.0    174.0                     0  Transient-Party    0.00   
2409      0.0    174.0                     0        Transient    0.00   
3181     38.0      0.0                     0  Transient-Party    0.00   
3684    308.0      0.0                   122  Transient-Party    0.00   
3708    308.0      0.0                   122  Transient-Party    0.00   
…        …      …                   …              …     …
115029    7.0      0.0                     0        Transient  100.80   
115091    0.0      0.0                     0        Transient    0.00   
116251  425.0      0.0                     0        Transient   73.80   
116534    9.0      0.0                     0  Transient-Party   22.86   
117087   52.0      0.0                     0        Transient    0.00   

        required_car_parking_spaces  total_of_special_requests  \
2224                              0                          0   
2409                              0                          0   
3181                              0                          0   
3684                              0                          0   
3708                              0                          0   
…                             …                         …
115029                            0                          0   
115091                            1                          1   
116251                            0                          0   
116534                            0                          1   
117087                            0                          0   

       reservation_status reservation_status_date  
2224            Check-Out               10/6/2015  
2409            Check-Out              10/12/2015  
3181            Check-Out              11/23/2015  
3684            Check-Out                1/4/2016  
3708            Check-Out                1/5/2016  
…                   …                     …
115029          Check-Out               6/30/2017  
```

```
115091          Check-Out               7/1/2017
116251          Check-Out              7/17/2017
116534          Check-Out              7/22/2017
117087          Check-Out              7/29/2017

[180 rows x 32 columns]
```

[64]:
```
#since we lacked some columns in the previous output, now we fix that
#with set_option in pamdas

pd.set_option('display.max_column',32)
```

[65]:
```
filter =(hotel["adults"]== 0) & (hotel["children"]==0) & (hotel['babies']==0)
hotel[filter]
```

[65]:
```
               hotel  is_canceled  lead_time  arrival_date_year  \
2224    Resort Hotel            0          1               2015
2409    Resort Hotel            0          0               2015
3181    Resort Hotel            0         36               2015
3684    Resort Hotel            0        165               2015
3708    Resort Hotel            0        165               2015
…                …          …        …               …
115029    City Hotel            0        107               2017
115091    City Hotel            0          1               2017
116251    City Hotel            0         44               2017
116534    City Hotel            0          2               2017
117087    City Hotel            0        170               2017

       arrival_date_month  arrival_date_week_number  \
2224              October                        41
2409              October                        42
3181             November                        47
3684             December                        53
3708             December                        53
…                    …                       …
115029               June                        26
115091               June                        26
116251               July                        28
116534               July                        28
117087               July                        30

        arrival_date_day_of_month  stays_in_weekend_nights  \
2224                            6                        0
2409                           12                        0
3181                           20                        1
3684                           30                        1
3708                           30                        2
```

```
…                           …                      …
115029                         27                        0
115091                         30                        0
116251                         15                        1
116534                         15                        2
117087                         27                        0

        stays_in_week_nights  adults  children  babies meal country  \
2224                       3       0       0.0       0   SC     PRT
2409                       0       0       0.0       0   SC     PRT
3181                       2       0       0.0       0   SC     ESP
3684                       4       0       0.0       0   SC     PRT
3708                       4       0       0.0       0   SC     PRT
…                      …     …       …     …  …     …
115029                     3       0       0.0       0   BB     CHE
115091                     1       0       0.0       0   SC     PRT
116251                     1       0       0.0       0   SC     SWE
116534                     5       0       0.0       0   SC     RUS
117087                     2       0       0.0       0   BB     BRA

       market_segment distribution_channel  is_repeated_guest  \
2224        Corporate            Corporate                  0
2409        Corporate            Corporate                  0
3181           Groups                TA/TO                  0
3684           Groups                TA/TO                  0
3708           Groups                TA/TO                  0
…               …                  …                …
115029      Online TA                TA/TO                  0
115091  Complementary               Direct                  0
116251      Online TA                TA/TO                  0
116534      Online TA                TA/TO                  0
117087  Offline TA/TO                TA/TO                  0

        previous_cancellations  previous_bookings_not_canceled  \
2224                         0                               0
2409                         0                               0
3181                         0                               0
3684                         0                               0
3708                         0                               0
…                        …                              …
115029                       0                               0
115091                       0                               0
116251                       0                               0
116534                       0                               0
117087                       0                               0

       reserved_room_type assigned_room_type  booking_changes deposit_type  \
2224                    A                  I                1   No Deposit
2409                    A                  I                0   No Deposit
3181                    A                  C                0   No Deposit
3684                    A                  A                1   No Deposit
3708                    A                  C                1   No Deposit
…                   …                 …              …          …
115029                  A                  A                1   No Deposit
115091                  E                  K                0   No Deposit
116251                  A                  K                2   No Deposit
116534                  A                  K                1   No Deposit
117087                  A                  A                0   No Deposit

        agent  company  days_in_waiting_list    customer_type     adr  \
2224      0.0    174.0                     0  Transient-Party    0.00
2409      0.0    174.0                     0        Transient    0.00
3181     38.0      0.0                     0  Transient-Party    0.00
3684    308.0      0.0                   122  Transient-Party    0.00
3708    308.0      0.0                   122  Transient-Party    0.00
…       …      …                   …              …     …
115029    7.0      0.0                     0        Transient  100.80
115091    0.0      0.0                     0        Transient    0.00
116251  425.0      0.0                     0        Transient   73.80
116534    9.0      0.0                     0  Transient-Party   22.86
117087   52.0      0.0                     0        Transient    0.00

        required_car_parking_spaces  total_of_special_requests  \
2224                              0                          0
2409                              0                          0
3181                              0                          0
3684                              0                          0
3708                              0                          0
…                              …                        …
115029                            0                          0
115091                            1                          1
116251                            0                          0
116534                            0                          1
117087                            0                          0

       reservation_status reservation_status_date
2224            Check-Out               10/6/2015
2409            Check-Out              10/12/2015
3181            Check-Out              11/23/2015
3684            Check-Out                1/4/2016
3708            Check-Out                1/5/2016
…                   …                      …
115029          Check-Out               6/30/2017
115091          Check-Out                7/1/2017
```

```
116251          Check-Out                 7/17/2017
116534          Check-Out                 7/22/2017
117087          Check-Out                 7/29/2017

[180 rows x 32 columns]
```

2 analysing the data

[66]:

```
#where the guests come from?
#spatial analysis

country_analysis = hotel[hotel['is_canceled']==0]['country'].value_counts().
 ↪reset_index()
```

[67]:

```
country_analysis
```

[67]:

```
    index  country
0     PRT    21071
1     GBR     9676
2     FRA     8481
3     ESP     6391
4     DEU     6069
..    …      …
161   BHR        1
162   AIA        1
163   BHS        1
164   TJK        1
165   BDI        1

[166 rows x 2 columns]
```

[68]:

```
#rename the columns

country_analysis.columns=['country','no.of guests']
```

[69]:

```
country_analysis
```

[69]:

```
    country  no.of guests
0       PRT         21071
1       GBR          9676
2       FRA          8481
3       ESP          6391
4       DEU          6069
..      …            …
161     BHR             1
162     AIA             1
163     BHS             1
```

```
164     TJK             1
165     BDI             1

[166 rows x 2 columns]
```

[70]:
```
!pip install folium
```

```
Requirement already satisfied: folium in c:\users\hp\anaconda3\lib\site-packages
(0.12.1)
Requirement already satisfied: branca>=0.3.0 in c:\users\hp\anaconda3\lib\site-
packages (from folium) (0.4.2)
Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\site-
packages (from folium) (2.24.0)
Requirement already satisfied: numpy in c:\users\hp\anaconda3\lib\site-packages
(from folium) (1.19.2)
Requirement already satisfied: jinja2>=2.9 in c:\users\hp\anaconda3\lib\site-
packages (from folium) (2.11.2)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
c:\users\hp\anaconda3\lib\site-packages (from requests->folium) (1.25.11)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\hp\anaconda3\lib\site-packages (from requests->folium) (2020.6.20)
Requirement already satisfied: chardet<4,>=3.0.2 in
c:\users\hp\anaconda3\lib\site-packages (from requests->folium) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in c:\users\hp\anaconda3\lib\site-
packages (from requests->folium) (2.10)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\hp\anaconda3\lib\site-packages (from jinja2>=2.9->folium) (1.1.1)
```

[71]:
```
import folium
from folium.plugins import HeatMap
```

[72]:
```
basemap = folium.Map()
```

[73]:
```
basemap
```

[73]:
```
<folium.folium.Map at 0x26b6622d670>
```

[74]:
```
!pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\hp\anaconda3\lib\site-packages
(4.14.3)
Requirement already satisfied: six in c:\users\hp\anaconda3\lib\site-packages
(from plotly) (1.15.0)
Requirement already satisfied: retrying>=1.3.3 in
c:\users\hp\anaconda3\lib\site-packages (from plotly) (1.3.3)
```

[75]:
```
import plotly.express as px
```

[76]:

```
map_guests = px.choropleth(country_analysis,
             locations = country_analysis['country'],
             color = country_analysis['no.of guests'],
             hover_name = country_analysis['country'],
             title='Home country of guests')
map_guests.show()
```

[77]:

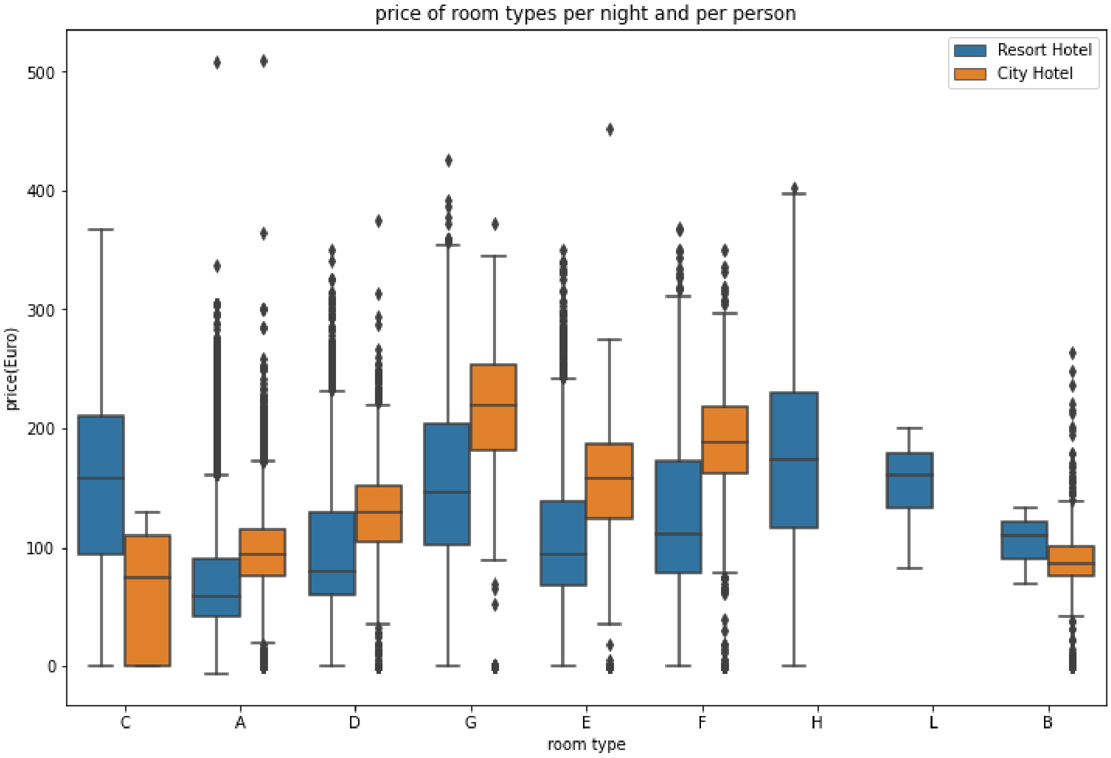
```
#how much guests pay for hotel
#distribution of hotel type

hotel_type = hotel[hotel['is_canceled']== 0]
hotel_type.columns
```

[77]:

```
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', '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', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

[78]:

```
plt.figure(figsize=(12,8))
sns.boxplot(x='reserved_room_type',y='adr',hue='hotel', data=hotel_type)
plt.title("price of room types per night and per person")
plt.xlabel('room type')
plt.ylabel('price(Euro)')
plt.legend()
plt.show()
```



```
[79]: # how does the price per night vary over the year
```

```
[80]: hotel.columns
```

```
[80]: 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', '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', 'total_of_special_requests',
              'reservation_status', 'reservation_status_date'],
             dtype='object')
```

```
[81]: resort_hotel=hotel[(hotel['hotel']=='Resort Hotel') & (hotel['is_canceled']==0)]
      city_hotel=hotel[(hotel['hotel']=='City Hotel') & (hotel['is_canceled']==0)]
```

```
[99]: resort_hotel.head()
```

```
[99]:          hotel  is_canceled  lead_time  arrival_date_year arrival_date_month  \
      0  Resort Hotel            0        342               2015               July
      1  Resort Hotel            0        737               2015               July
      2  Resort Hotel            0          7               2015               July
      3  Resort Hotel            0         13               2015               July
      4  Resort Hotel            0         14               2015               July

         arrival_date_week_number  arrival_date_day_of_month  \
      0                        27                          1
      1                        27                          1
      2                        27                          1
      3                        27                          1
      4                        27                          1

         stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
      0                        0                     0       2       0.0       0
      1                        0                     0       2       0.0       0
      2                        0                     1       1       0.0       0
      3                        0                     1       1       0.0       0
      4                        0                     2       2       0.0       0

        meal country market_segment distribution_channel  is_repeated_guest  \
      0   BB     PRT         Direct               Direct                  0
      1   BB     PRT         Direct               Direct                  0
      2   BB     GBR         Direct               Direct                  0
      3   BB     GBR      Corporate            Corporate                  0
      4   BB     GBR      Online TA                TA/TO                  0

         previous_cancellations  previous_bookings_not_canceled reserved_room_type  \
      0                       0                               0                  C
      1                       0                               0                  C
      2                       0                               0                  A
      3                       0                               0                  A
      4                       0                               0                  A

        assigned_room_type  booking_changes deposit_type  agent  company  \
      0                  C                3   No Deposit    0.0      0.0
      1                  C                4   No Deposit    0.0      0.0
      2                  C                0   No Deposit    0.0      0.0
      3                  A                0   No Deposit  304.0      0.0
      4                  A                0   No Deposit  240.0      0.0

         days_in_waiting_list customer_type   adr  required_car_parking_spaces  \
      0                     0     Transient   0.0                            0
      1                     0     Transient   0.0                            0
      2                     0     Transient  75.0                            0
      3                     0     Transient  75.0                            0
```

```
4                      0     Transient   98.0                           0

   total_of_special_requests reservation_status reservation_status_date
0                          0          Check-Out                7/1/2015
1                          0          Check-Out                7/1/2015
2                          0          Check-Out                7/2/2015
3                          0          Check-Out                7/2/2015
4                          1          Check-Out                7/3/2015
```

[100]:

```
#group by arrival date month and price
#to make it a df, reset index

resort_hoteldf = resort_hotel.groupby(['arrival_date_month'])['adr'].mean().
 ↪reset_index()
```

[101]:

```
resort_hoteldf
```

[101]:

```
   arrival_date_month         adr
0               April   75.867816
1              August  181.205892
2            December   68.322236
3            February   54.147478
4             January   48.708919
5                July  150.122528
6                June  107.921869
7               March   57.012487
8                 May   76.657558
9            November   48.681640
10            October   61.727505
11          September   96.416860
```

[102]:

```
city_hoteldf = city_hotel.groupby(['arrival_date_month'])['adr'].mean().
 ↪reset_index()
```

[103]:

```
city_hoteldf
```

[103]:

```
   arrival_date_month         adr
0               April  111.856824
1              August  118.412083
2            December   87.856764
3            February   86.183025
4             January   82.160634
5                July  115.563810
6                June  117.702075
7               March   90.170722
8                 May  120.445842
9            November   86.500456
```

```
10            October  101.745956
11          September  112.598452
```

[104]:

```
#merge the two dataframes

final = resort_hoteldf.merge(city_hoteldf, on='arrival_date_month')
final.columns = ['months','price_for_resort','price_for_city']
```

[105]:

```
final
```

[105]:

```
       months  price_for_resort  price_for_city
0       April         75.867816      111.856824
1      August        181.205892      118.412083
2    December         68.322236       87.856764
3    February         54.147478       86.183025
4     January         48.708919       82.160634
5        July        150.122528      115.563810
6        June        107.921869      117.702075
7       March         57.012487       90.170722
8         May         76.657558      120.445842
9    November         48.681640       86.500456
10    October         61.727505      101.745956
11  September         96.416860      112.598452
```

[106]:

```
#sort the months using

!pip install sorted-months-weekdays
```

```
Requirement already satisfied: sorted-months-weekdays in
c:\users\hp\anaconda3\lib\site-packages (0.2)
```

[107]:

```
!pip install sort-dataframeby-monthorweek
```

```
Requirement already satisfied: sort-dataframeby-monthorweek in
c:\users\hp\anaconda3\lib\site-packages (0.4)
```

[108]:

```
import sort_dataframeby_monthorweek as sd
```

[109]:

```
def sort_data(df,colname):
    return sd.Sort_Dataframeby_Month(df,colname)
```

[110]:

```
final = sort_data(final,'months')
final
```

[110]:

```
       months  price_for_resort  price_for_city
0     January         48.708919       82.160634
1    February         54.147478       86.183025
2       March         57.012487       90.170722
```

```
3       April        75.867816      111.856824
4         May        76.657558      120.445842
5        June       107.921869      117.702075
6        July       150.122528      115.563810
7      August       181.205892      118.412083
8   September        96.416860      112.598452
9     October        61.727505      101.745956
10   November        48.681640       86.500456
11   December        68.322236       87.856764
```

```
[111]: #visuals.
       #line plot

       px.line(final,x='months',y=['price_for_resort', 'price_for_city'],
              title='room price overnight per month')
```

```
[112]: final.columns
```

```
[112]: Index(['months', 'price_for_resort', 'price_for_city'], dtype='object')
```

```
[115]: #analysis demands of hotels
```

```
[116]: rush_resort = resort_hotel['arrival_date_month'].value_counts().reset_index()
       rush_resort.columns = ['months','no.of guests']
       rush_resort
```

```
[116]:        months  no.of guests
       0      August          3257
       1        July          3137
       2     October          2577
       3       March          2573
       4       April          2550
       5         May          2535
       6    February          2308
       7   September          2102
       8        June          2038
       9    December          2017
       10   November          1976
       11    January          1868
```

```
[117]: rush_city = city_hotel['arrival_date_month'].value_counts().reset_index()
       rush_city.columns = ['months','no.of guests']
       rush_city
```

```
[117]:        months  no.of guests
       0      August          5381
       1        July          4782
       2         May          4579
```

```
3         June          4366
4      October          4337
5    September          4290
6        March          4072
7        April          4015
8     February          3064
9     November          2696
10    December          2392
11     January          2254
```

```
[118]: #merge dataframes

       final_rush = rush_resort.merge(rush_city,on='months')
       final_rush.columns = ['months','no of guests in resort','no of guest in city']
       final_rush
```

```
[118]:        months  no of guests in resort  no of guest in city
       0      August                    3257                 5381
       1        July                    3137                 4782
       2     October                    2577                 4337
       3       March                    2573                 4072
       4       April                    2550                 4015
       5         May                    2535                 4579
       6    February                    2308                 3064
       7   September                    2102                 4290
       8        June                    2038                 4366
       9    December                    2017                 2392
       10   November                    1976                 2696
       11    January                    1868                 2254
```

```
[119]: #hierachy of my months
       final_rush = sort_data(final_rush,'months')
       final_rush
```

```
[119]:        months  no of guests in resort  no of guest in city
       0     January                    1868                 2254
       1    February                    2308                 3064
       2       March                    2573                 4072
       3       April                    2550                 4015
       4         May                    2535                 4579
       5        June                    2038                 4366
       6        July                    3137                 4782
       7      August                    3257                 5381
       8   September                    2102                 4290
       9     October                    2577                 4337
       10   November                    1976                 2696
       11   December                    2017                 2392
```

[120]:
```
#we need trend, so we go for line plot

px.line(final_rush,x='months',y= ['no of guests in resort', 'no of guest in␣
↪city'],
       title='total no of guest per months')
```

3 machine learning

[121]:
```
hotel.head()
```

[121]:
```
          hotel  is_canceled  lead_time  arrival_date_year arrival_date_month  \
0  Resort Hotel            0        342               2015               July
1  Resort Hotel            0        737               2015               July
2  Resort Hotel            0          7               2015               July
3  Resort Hotel            0         13               2015               July
4  Resort Hotel            0         14               2015               July

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1
1                        27                          1
2                        27                          1
3                        27                          1
4                        27                          1

   stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
0                        0                     0       2       0.0       0
1                        0                     0       2       0.0       0
2                        0                     1       1       0.0       0
3                        0                     1       1       0.0       0
4                        0                     2       2       0.0       0

  meal country market_segment distribution_channel  is_repeated_guest  \
0   BB     PRT         Direct               Direct                  0
1   BB     PRT         Direct               Direct                  0
2   BB     GBR         Direct               Direct                  0
3   BB     GBR      Corporate            Corporate                  0
4   BB     GBR      Online TA                TA/TO                  0

   previous_cancellations  previous_bookings_not_canceled reserved_room_type  \
0                       0                               0                  C
1                       0                               0                  C
2                       0                               0                  A
3                       0                               0                  A
4                       0                               0                  A

  assigned_room_type  booking_changes deposit_type  agent  company  \
0                  C                3   No Deposit    0.0      0.0
1                  C                4   No Deposit    0.0      0.0
2                  C                0   No Deposit    0.0      0.0
3                  A                0   No Deposit  304.0      0.0
4                  A                0   No Deposit  240.0      0.0

   days_in_waiting_list customer_type   adr  required_car_parking_spaces  \
0                     0     Transient   0.0                            0
1                     0     Transient   0.0                            0
2                     0     Transient  75.0                            0
3                     0     Transient  75.0                            0
4                     0     Transient  98.0                            0

   total_of_special_requests reservation_status reservation_status_date
0                          0          Check-Out                7/1/2015
1                          0          Check-Out                7/1/2015
2                          0          Check-Out                7/2/2015
3                          0          Check-Out                7/2/2015
4                          1          Check-Out                7/3/2015
```

[122]:
```
#find correlation
hotel.corr()
```

[122]:
```
                                is_canceled  lead_time  arrival_date_year  \
is_canceled                        1.000000   0.293123           0.016660
lead_time                          0.293123   1.000000           0.040142
arrival_date_year                  0.016660   0.040142           1.000000
arrival_date_week_number           0.008148   0.126871          -0.540561
arrival_date_day_of_month         -0.006130   0.002268          -0.000221
stays_in_weekend_nights           -0.001791   0.085671           0.021497
stays_in_week_nights               0.024765   0.165799           0.030883
adults                             0.060017   0.119519           0.029635
children                           0.005036  -0.037613           0.054636
babies                            -0.032491  -0.020915          -0.013192
is_repeated_guest                 -0.084793  -0.124410           0.010341
previous_cancellations             0.110133   0.086042          -0.119822
previous_bookings_not_canceled    -0.057358  -0.073548           0.029218
booking_changes                   -0.144381   0.000149           0.030872
agent                             -0.046529  -0.012640           0.056463
company                           -0.082995  -0.086250           0.033882
days_in_waiting_list               0.054186   0.170084          -0.056497
adr                                0.047557  -0.063077           0.197580
required_car_parking_spaces       -0.195498  -0.116451          -0.013684
total_of_special_requests         -0.234658  -0.095712           0.108531

                                arrival_date_week_number  \
is_canceled                                     0.008148
```

```
lead_time                                       0.126871
arrival_date_year                              -0.540561
arrival_date_week_number                        1.000000
arrival_date_day_of_month                       0.066809
stays_in_weekend_nights                         0.018208
stays_in_week_nights                            0.015558
adults                                          0.025909
children                                        0.005515
babies                                          0.010395
is_repeated_guest                              -0.030131
previous_cancellations                          0.035501
previous_bookings_not_canceled                 -0.020904
booking_changes                                 0.005508
agent                                          -0.018244
company                                        -0.032750
days_in_waiting_list                            0.022933
adr                                             0.075791
required_car_parking_spaces                     0.001920
total_of_special_requests                       0.026149

                                arrival_date_day_of_month  \
is_canceled                                     -0.006130
lead_time                                        0.002268
arrival_date_year                               -0.000221
arrival_date_week_number                         0.066809
arrival_date_day_of_month                        1.000000
stays_in_weekend_nights                         -0.016354
stays_in_week_nights                            -0.028174
adults                                          -0.001566
children                                         0.014553
babies                                          -0.000230
is_repeated_guest                               -0.006145
previous_cancellations                          -0.027011
previous_bookings_not_canceled                  -0.000300
booking_changes                                  0.010613
agent                                            0.000202
company                                          0.003724
days_in_waiting_list                             0.022728
adr                                              0.030245
required_car_parking_spaces                      0.008683
total_of_special_requests                        0.003062

                                stays_in_weekend_nights  stays_in_week_nights  \
is_canceled                                   -0.001791              0.024765
lead_time                                      0.085671              0.165799
arrival_date_year                              0.021497              0.030883
arrival_date_week_number                       0.018208              0.015558
```

```
arrival_date_day_of_month                     -0.016354             -0.028174
stays_in_weekend_nights                        1.000000              0.498969
stays_in_week_nights                           0.498969              1.000000
adults                                         0.091871              0.092976
children                                       0.045794              0.044203
babies                                         0.018483              0.020191
is_repeated_guest                             -0.087239             -0.097245
previous_cancellations                        -0.012775             -0.013992
previous_bookings_not_canceled                -0.042715             -0.048743
booking_changes                                0.063281              0.096209
agent                                          0.161427              0.195135
company                                       -0.079977             -0.043641
days_in_waiting_list                          -0.054151             -0.002020
adr                                            0.049342              0.065237
required_car_parking_spaces                   -0.018554             -0.024859
total_of_special_requests                      0.072671              0.068192

                                  adults  children    babies  \
is_canceled                     0.060017  0.005036 -0.032491
lead_time                       0.119519 -0.037613 -0.020915
arrival_date_year               0.029635  0.054636 -0.013192
arrival_date_week_number        0.025909  0.005515  0.010395
arrival_date_day_of_month      -0.001566  0.014553 -0.000230
stays_in_weekend_nights         0.091871  0.045794  0.018483
stays_in_week_nights            0.092976  0.044203  0.020191
adults                          1.000000  0.030440  0.018146
children                        0.030440  1.000000  0.024030
babies                          0.018146  0.024030  1.000000
is_repeated_guest              -0.146426 -0.032858 -0.008943
previous_cancellations         -0.006738 -0.024729 -0.007501
previous_bookings_not_canceled -0.107983 -0.021072 -0.006550
booking_changes                -0.051673  0.048952  0.083440
agent                           0.024994  0.050581  0.030266
company                        -0.166778 -0.042622 -0.009459
days_in_waiting_list           -0.008283 -0.033271 -0.010621
adr                             0.230641  0.324853  0.029186
required_car_parking_spaces     0.014785  0.056255  0.037383
total_of_special_requests       0.122884  0.081736  0.097889

                                is_repeated_guest  previous_cancellations  \
is_canceled                             -0.084793                0.110133
lead_time                               -0.124410                0.086042
arrival_date_year                        0.010341               -0.119822
arrival_date_week_number                -0.030131                0.035501
arrival_date_day_of_month               -0.006145               -0.027011
stays_in_weekend_nights                 -0.087239               -0.012775
stays_in_week_nights                    -0.097245               -0.013992
```

```
adults                                  -0.146426               -0.006738
children                                -0.032858               -0.024729
babies                                  -0.008943               -0.007501
is_repeated_guest                        1.000000                0.082293
previous_cancellations                   0.082293                1.000000
previous_bookings_not_canceled           0.418056                0.152728
booking_changes                          0.012092               -0.026993
agent                                   -0.052264               -0.018192
company                                  0.159723               -0.001190
days_in_waiting_list                    -0.022235                0.005929
adr                                     -0.134314               -0.065646
required_car_parking_spaces              0.077090               -0.018492
total_of_special_requests                0.013050               -0.048384

                                previous_bookings_not_canceled  \
is_canceled                                          -0.057358
lead_time                                            -0.073548
arrival_date_year                                     0.029218
arrival_date_week_number                             -0.020904
arrival_date_day_of_month                            -0.000300
stays_in_weekend_nights                              -0.042715
stays_in_week_nights                                 -0.048743
adults                                               -0.107983
children                                             -0.021072
babies                                               -0.006550
is_repeated_guest                                     0.418056
previous_cancellations                                0.152728
previous_bookings_not_canceled                        1.000000
booking_changes                                       0.011608
agent                                                -0.046296
company                                               0.110817
days_in_waiting_list                                 -0.009397
adr                                                  -0.072144
required_car_parking_spaces                           0.047653
total_of_special_requests                             0.037824

                                booking_changes     agent   company  \
is_canceled                           -0.144381 -0.046529 -0.082995
lead_time                              0.000149 -0.012640 -0.086250
arrival_date_year                      0.030872  0.056463  0.033882
arrival_date_week_number               0.005508 -0.018244 -0.032750
arrival_date_day_of_month              0.010613  0.000202  0.003724
stays_in_weekend_nights                0.063281  0.161427 -0.079977
stays_in_week_nights                   0.096209  0.195135 -0.043641
adults                                -0.051673  0.024994 -0.166778
children                               0.048952  0.050581 -0.042622
babies                                 0.083440  0.030266 -0.009459
```

```
is_repeated_guest                      0.012092 -0.052264  0.159723
previous_cancellations                -0.026993 -0.018192 -0.001190
previous_bookings_not_canceled         0.011608 -0.046296  0.110817
booking_changes                        1.000000  0.036478  0.088863
agent                                  0.036478  1.000000 -0.121536
company                                0.088863 -0.121536  1.000000
days_in_waiting_list                  -0.011634 -0.040853 -0.022986
adr                                    0.019618  0.016707 -0.128470
required_car_parking_spaces            0.065620  0.119158  0.038299
total_of_special_requests              0.052833  0.060696 -0.091066

                                days_in_waiting_list       adr  \
is_canceled                                 0.054186  0.047557
lead_time                                   0.170084 -0.063077
arrival_date_year                          -0.056497  0.197580
arrival_date_week_number                    0.022933  0.075791
arrival_date_day_of_month                   0.022728  0.030245
stays_in_weekend_nights                    -0.054151  0.049342
stays_in_week_nights                       -0.002020  0.065237
adults                                     -0.008283  0.230641
children                                   -0.033271  0.324853
babies                                     -0.010621  0.029186
is_repeated_guest                          -0.022235 -0.134314
previous_cancellations                      0.005929 -0.065646
previous_bookings_not_canceled             -0.009397 -0.072144
booking_changes                            -0.011634  0.019618
agent                                      -0.040853  0.016707
company                                    -0.022986 -0.128470
days_in_waiting_list                        1.000000 -0.040756
adr                                        -0.040756  1.000000
required_car_parking_spaces                -0.030600  0.056628
total_of_special_requests                  -0.082730  0.172185

                                required_car_parking_spaces  \
is_canceled                                       -0.195498
lead_time                                         -0.116451
arrival_date_year                                 -0.013684
arrival_date_week_number                           0.001920
arrival_date_day_of_month                          0.008683
stays_in_weekend_nights                           -0.018554
stays_in_week_nights                              -0.024859
adults                                             0.014785
children                                           0.056255
babies                                             0.037383
is_repeated_guest                                  0.077090
previous_cancellations                            -0.018492
previous_bookings_not_canceled                     0.047653
```

```
booking_changes                                     0.065620
agent                                               0.119158
company                                             0.038299
days_in_waiting_list                               -0.030600
adr                                                 0.056628
required_car_parking_spaces                         1.000000
total_of_special_requests                           0.082626

                                total_of_special_requests
is_canceled                                     -0.234658
lead_time                                       -0.095712
arrival_date_year                                0.108531
arrival_date_week_number                         0.026149
arrival_date_day_of_month                        0.003062
stays_in_weekend_nights                          0.072671
stays_in_week_nights                             0.068192
adults                                           0.122884
children                                         0.081736
babies                                           0.097889
is_repeated_guest                                0.013050
previous_cancellations                          -0.048384
previous_bookings_not_canceled                   0.037824
booking_changes                                  0.052833
agent                                            0.060696
company                                         -0.091066
days_in_waiting_list                            -0.082730
adr                                              0.172185
required_car_parking_spaces                      0.082626
total_of_special_requests                        1.000000
```

[123]:

```
#correlation with respet to is cancelled

co_relate = hotel.corr()['is_canceled']
co_relate
```

[123]:

```
is_canceled                       1.000000
lead_time                         0.293123
arrival_date_year                 0.016660
arrival_date_week_number          0.008148
arrival_date_day_of_month        -0.006130
stays_in_weekend_nights          -0.001791
stays_in_week_nights              0.024765
adults                            0.060017
children                          0.005036
babies                           -0.032491
is_repeated_guest                -0.084793
previous_cancellations            0.110133
```

```
previous_bookings_not_canceled   -0.057358
booking_changes                  -0.144381
agent                            -0.046529
company                          -0.082995
days_in_waiting_list              0.054186
adr                               0.047557
required_car_parking_spaces      -0.195498
total_of_special_requests        -0.234658
Name: is_canceled, dtype: float64
```

[124]:

```
#finding the most important features

co_relate.abs().sort_values(ascending=False)
```

[124]:

```
is_canceled                       1.000000
lead_time                         0.293123
total_of_special_requests         0.234658
required_car_parking_spaces       0.195498
booking_changes                   0.144381
previous_cancellations            0.110133
is_repeated_guest                 0.084793
company                           0.082995
adults                            0.060017
previous_bookings_not_canceled    0.057358
days_in_waiting_list              0.054186
adr                               0.047557
agent                             0.046529
babies                            0.032491
stays_in_week_nights              0.024765
arrival_date_year                 0.016660
arrival_date_week_number          0.008148
arrival_date_day_of_month         0.006130
children                          0.005036
stays_in_weekend_nights           0.001791
Name: is_canceled, dtype: float64
```

[125]:

```
#
hotel.groupby('is_canceled')['reservation_status'].value_counts()
```

[125]:

```
is_canceled  reservation_status
0            Check-Out             75166
1            Canceled              43017
             No-Show                1207
Name: reservation_status, dtype: int64
```

[126]:

```
#exclude unnecessary features
```

```
list_not = ['days_in_waiting_list ','arrival_date_year ']
```

[127]:

```
#fetch numerical features we have
#using a list comprehension

num_features = [col for col in hotel.columns if hotel[col].dtype != 'object'␣
 ↪and col not in list_not]
num_features
```

[127]:

```
['is_canceled',
 'lead_time',
 'arrival_date_year',
 'arrival_date_week_number',
 'arrival_date_day_of_month',
 'stays_in_weekend_nights',
 'stays_in_week_nights',
 'adults',
 'children',
 'babies',
 'is_repeated_guest',
 'previous_cancellations',
 'previous_bookings_not_canceled',
 'booking_changes',
 'agent',
 'company',
 'days_in_waiting_list',
 'adr',
 'required_car_parking_spaces',
 'total_of_special_requests']
```

[128]:

```
hotel.columns
```

[128]:

```
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', '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', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

[129]:

```
cat_not =␣
↪['arrival_date_year','country','assigned_room_type','booking_changes',␣
↪'reservation_status','days_in_waiting_list']
```

[130]:
```
cat_features = [col for col in hotel.columns if hotel[col].dtype == 'object'␣
↪and col not in cat_not]
```

[131]:
```
cat_features
```

[131]:
```
['hotel',
 'arrival_date_month',
 'meal',
 'market_segment',
 'distribution_channel',
 'reserved_room_type',
 'deposit_type',
 'customer_type',
 'reservation_status_date']
```

[132]:
```
#extracting derived features
```

[133]:
```
cat_data = hotel[cat_features]
```

[134]:
```
cat_data.dtypes
```

[134]:
```
hotel                          object
arrival_date_month             object
meal                           object
market_segment                 object
distribution_channel           object
reserved_room_type             object
deposit_type                   object
customer_type                  object
reservation_status_date        object
dtype: object
```

[135]:
```
#when you want to block the warning
import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

[136]:
```
cat_data['reservation_status_date'] = pd.
↪to_datetime(cat_data['reservation_status_date'])
```

[137]:
```
#creating different columns for month day and year

cat_data['year'] = cat_data['reservation_status_date'].dt.year
cat_data['month'] = cat_data['reservation_status_date'].dt.month
```

```
cat_data['day'] = cat_data['reservation_status_date'].dt.day
```

[138]:
```
#drop the column with the combination of the data

cat_data.drop('reservation_status_date', axis =1, inplace=True)
```

[139]:
```
cat_data.dtypes
```

[139]:
```
hotel                   object
arrival_date_month      object
meal                    object
market_segment          object
distribution_channel    object
reserved_room_type      object
deposit_type            object
customer_type           object
year                     int64
month                    int64
day                      int64
dtype: object
```

[140]:
```
cat_data['cancellation']=hotel['is_canceled']
```

[ ]:
```

```

[141]:
```
cat_data.dtypes
```

[141]:
```
hotel                   object
arrival_date_month      object
meal                    object
market_segment          object
distribution_channel    object
reserved_room_type      object
deposit_type            object
customer_type           object
year                     int64
month                    int64
day                      int64
cancellation             int64
dtype: object
```

[142]:
```
#applying feature encoding
cat_data.head()
```

[142]:
```
          hotel arrival_date_month meal market_segment distribution_channel  \
0  Resort Hotel               July   BB         Direct               Direct
1  Resort Hotel               July   BB         Direct               Direct
2  Resort Hotel               July   BB         Direct               Direct
```

```
3  Resort Hotel                July   BB      Corporate            Corporate
4  Resort Hotel                July   BB      Online TA                TA/TO

  reserved_room_type deposit_type customer_type  year  month  day  \
0                  C   No Deposit     Transient  2015      7    1
1                  C   No Deposit     Transient  2015      7    1
2                  A   No Deposit     Transient  2015      7    2
3                  A   No Deposit     Transient  2015      7    2
4                  A   No Deposit     Transient  2015      7    3

   cancellation
0             0
1             0
2             0
3             0
4             0
```

[143]:
```
##mean encoding technique

cat_data['market_segment'].unique()
```

[143]:
```
array(['Direct', 'Corporate', 'Online TA', 'Offline TA/TO',
       'Complementary', 'Groups', 'Undefined', 'Aviation'], dtype=object)
```

[144]:
```
col_enc = cat_data.columns[0:8]
```

[145]:
```
for col in col_enc:
    print(cat_data.groupby([col])['cancellation'].mean().to_dict())
    print('\n')
```

```
{'City Hotel': 0.41726963317786464, 'Resort Hotel': 0.27763354967548676}


{'April': 0.4079718640093787, 'August': 0.3775311666786769, 'December':
0.3497050147492625, 'February': 0.3341596430342092, 'January':
0.3047731489289931, 'July': 0.37453597662112, 'June': 0.4145717158789652,
'March': 0.3215233816622422, 'May': 0.39665846832329743, 'November':
0.3123344127171033, 'October': 0.3804659498207885, 'September':
0.3917015607156452}


{'BB': 0.3738489871086556, 'FB': 0.5989974937343359, 'HB': 0.3446034709258107,
'SC': 0.3723943661971831, 'Undefined': 0.2446535500427716}


{'Aviation': 0.21940928270042195, 'Complementary': 0.13055181695827725,
'Corporate': 0.1873465533522191, 'Direct': 0.15341900682214818, 'Groups':
0.6106203624249155, 'Offline TA/TO': 0.34316032866757507, 'Online TA':
```

```
0.3672114312020823, 'Undefined': 1.0}


{'Corporate': 0.22075782537067545, 'Direct': 0.17459883919426425, 'GDS':
0.19170984455958548, 'TA/TO': 0.41025850618166954, 'Undefined': 0.8}


{'A': 0.3910737958462218, 'B': 0.3291592128801431, 'C': 0.33047210300429186,
'D': 0.3177959481277017, 'E': 0.29288446824789593, 'F': 0.30376251294442524,
'G': 0.3643744030563515, 'H': 0.40765391014975044, 'L': 0.3333333333333333, 'P':
1.0}


{'No Deposit': 0.28377022390841067, 'Non Refund': 0.9936244601357374,
'Refundable': 0.2222222222222222}


{'Contract': 0.3096172718351325, 'Group': 0.1022530329289428, 'Transient':
0.4074632028835102, 'Transient-Party': 0.2542986785543703}
```

[146]:

```
for col in col_enc:
    dict = cat_data.groupby([col])['cancellation'].mean().to_dict()
    cat_data[col] = cat_data[col].map(dict)
```

[147]:

```
cat_data.head()
```

[147]:

```
      hotel  arrival_date_month      meal  market_segment  \
0  0.277634            0.374536  0.373849        0.153419
1  0.277634            0.374536  0.373849        0.153419
2  0.277634            0.374536  0.373849        0.153419
3  0.277634            0.374536  0.373849        0.187347
4  0.277634            0.374536  0.373849        0.367211

   distribution_channel  reserved_room_type  deposit_type  customer_type  \
0              0.174599            0.330472       0.28377       0.407463
1              0.174599            0.330472       0.28377       0.407463
2              0.174599            0.391074       0.28377       0.407463
3              0.220758            0.391074       0.28377       0.407463
4              0.410259            0.391074       0.28377       0.407463

   year  month  day  cancellation
0  2015      7    1             0
1  2015      7    1             0
2  2015      7    2             0
3  2015      7    2             0
```

```
4  2015      7    3             0
```

[148]:
```
num_features
```

[148]:
```
['is_canceled',
 'lead_time',
 'arrival_date_year',
 'arrival_date_week_number',
 'arrival_date_day_of_month',
 'stays_in_weekend_nights',
 'stays_in_week_nights',
 'adults',
 'children',
 'babies',
 'is_repeated_guest',
 'previous_cancellations',
 'previous_bookings_not_canceled',
 'booking_changes',
 'agent',
 'company',
 'days_in_waiting_list',
 'adr',
 'required_car_parking_spaces',
 'total_of_special_requests']
```

[149]:
```
entire_df = pd.concat([cat_data,hotel[num_features]], axis=1)
```

[150]:
```
entire_df.head()
```

[150]:
```
      hotel  arrival_date_month      meal  market_segment  \
0  0.277634            0.374536  0.373849        0.153419
1  0.277634            0.374536  0.373849        0.153419
2  0.277634            0.374536  0.373849        0.153419
3  0.277634            0.374536  0.373849        0.187347
4  0.277634            0.374536  0.373849        0.367211

   distribution_channel  reserved_room_type  deposit_type  customer_type  \
0              0.174599            0.330472       0.28377       0.407463
1              0.174599            0.330472       0.28377       0.407463
2              0.174599            0.391074       0.28377       0.407463
3              0.220758            0.391074       0.28377       0.407463
4              0.410259            0.391074       0.28377       0.407463

   year  month  day  cancellation  is_canceled  lead_time  arrival_date_year  \
0  2015      7    1             0            0        342               2015
1  2015      7    1             0            0        737               2015
2  2015      7    2             0            0          7               2015
```

```
3  2015      7    2             0            0         13               2015
4  2015      7    3             0            0         14               2015

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1
1                        27                          1
2                        27                          1
3                        27                          1
4                        27                          1

   stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
0                        0                     0       2       0.0       0
1                        0                     0       2       0.0       0
2                        0                     1       1       0.0       0
3                        0                     1       1       0.0       0
4                        0                     2       2       0.0       0

   is_repeated_guest  previous_cancellations  previous_bookings_not_canceled  \
0                  0                       0                               0
1                  0                       0                               0
2                  0                       0                               0
3                  0                       0                               0
4                  0                       0                               0

   booking_changes  agent  company  days_in_waiting_list   adr  \
0                3    0.0      0.0                     0   0.0
1                4    0.0      0.0                     0   0.0
2                0    0.0      0.0                     0  75.0
3                0  304.0      0.0                     0  75.0
4                0  240.0      0.0                     0  98.0

   required_car_parking_spaces  total_of_special_requests
0                            0                          0
1                            0                          0
2                            0                          0
3                            0                          0
4                            0                          1
```

[151]:

```
#
entire_df.drop('cancellation', axis=1, inplace=True)
```

[152]:

```
entire_df.shape
```

[152]:

```
(119390, 31)
```

[153]:

```
#handling outliers
```

```
entire_df.head()
```

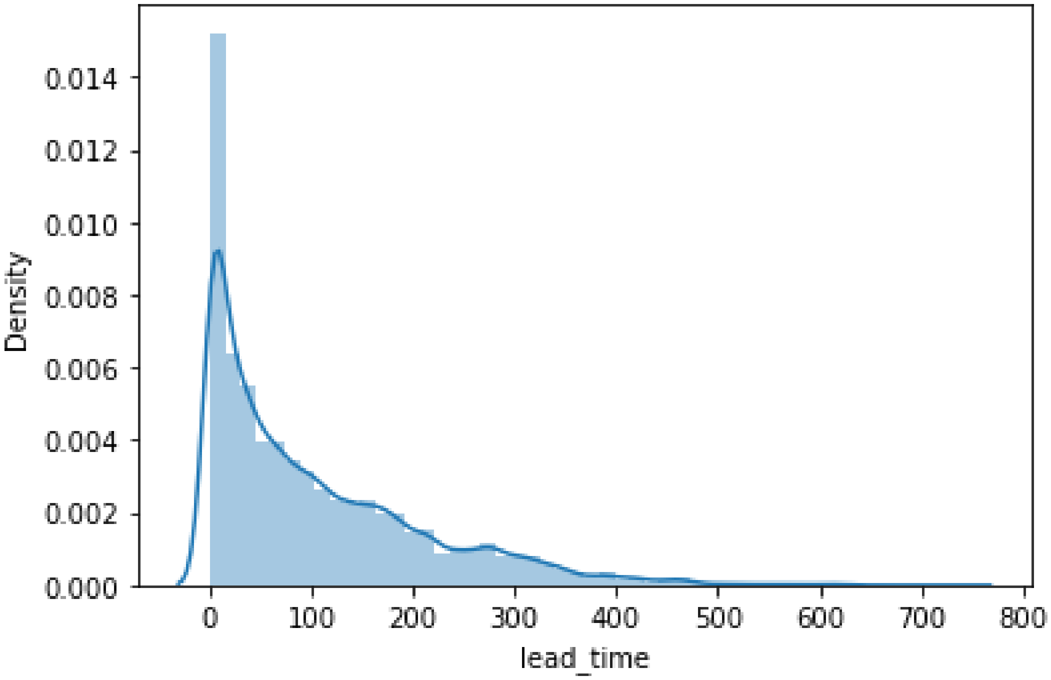
[153]:

```
      hotel  arrival_date_month      meal  market_segment  \
0  0.277634            0.374536  0.373849        0.153419   
1  0.277634            0.374536  0.373849        0.153419   
2  0.277634            0.374536  0.373849        0.153419   
3  0.277634            0.374536  0.373849        0.187347   
4  0.277634            0.374536  0.373849        0.367211   

   distribution_channel  reserved_room_type  deposit_type  customer_type  \
0              0.174599            0.330472       0.28377       0.407463   
1              0.174599            0.330472       0.28377       0.407463   
2              0.174599            0.391074       0.28377       0.407463   
3              0.220758            0.391074       0.28377       0.407463   
4              0.410259            0.391074       0.28377       0.407463   

   year  month  day  is_canceled  lead_time  arrival_date_year  \
0  2015      7    1            0        342               2015   
1  2015      7    1            0        737               2015   
2  2015      7    2            0          7               2015   
3  2015      7    2            0         13               2015   
4  2015      7    3            0         14               2015   

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1   
1                        27                          1   
2                        27                          1   
3                        27                          1   
4                        27                          1   

   stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
0                        0                     0       2       0.0       0   
1                        0                     0       2       0.0       0   
2                        0                     1       1       0.0       0   
3                        0                     1       1       0.0       0   
4                        0                     2       2       0.0       0   

   is_repeated_guest  previous_cancellations  previous_bookings_not_canceled  \
0                  0                       0                               0   
1                  0                       0                               0   
2                  0                       0                               0   
3                  0                       0                               0   
4                  0                       0                               0   

   booking_changes  agent  company  days_in_waiting_list  adr  \
0                3    0.0      0.0                     0  0.0   
1                4    0.0      0.0                     0  0.0   
```

```
2                0    0.0      0.0                     0  75.0
3                0  304.0      0.0                     0  75.0
4                0  240.0      0.0                     0  98.0

   required_car_parking_spaces  total_of_special_requests
0                            0                          0
1                            0                          0
2                            0                          0
3                            0                          0
4                            0                          1
```

[154]:
```
#distribution of lead time

sns.distplot(entire_df['lead_time'])
```

[154]:
```
<AxesSubplot:xlabel='lead_time', ylabel='Density'>
```

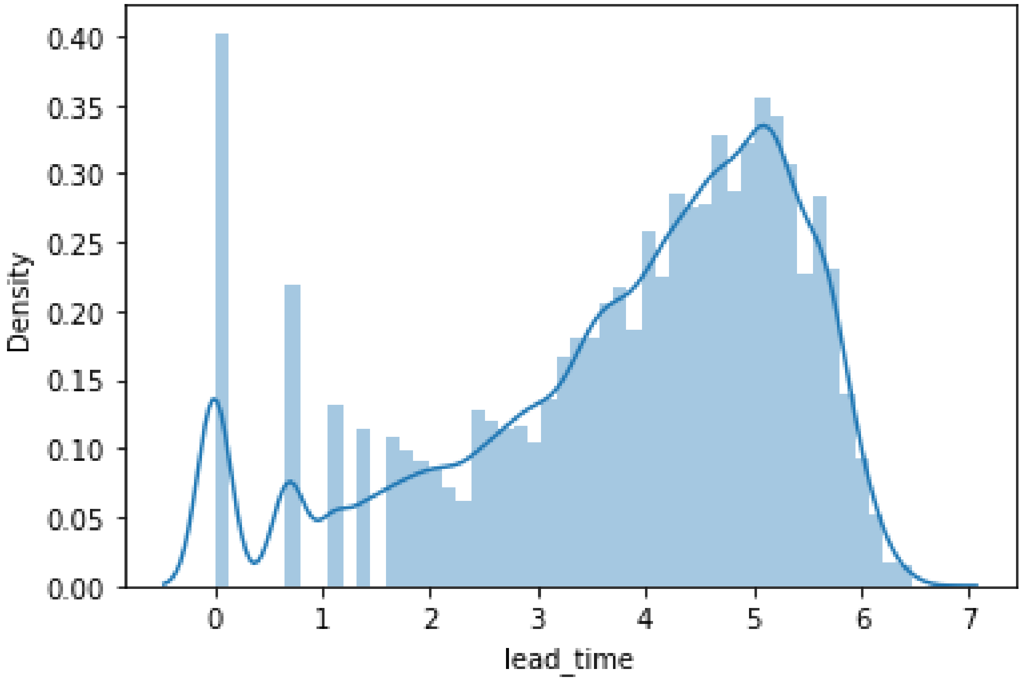


[155]:
```
#find the log of these

def handle_outlier(col):
    entire_df[col] = np.log1p(entire_df[col])
```

[156]:
```
handle_outlier('lead_time')
```

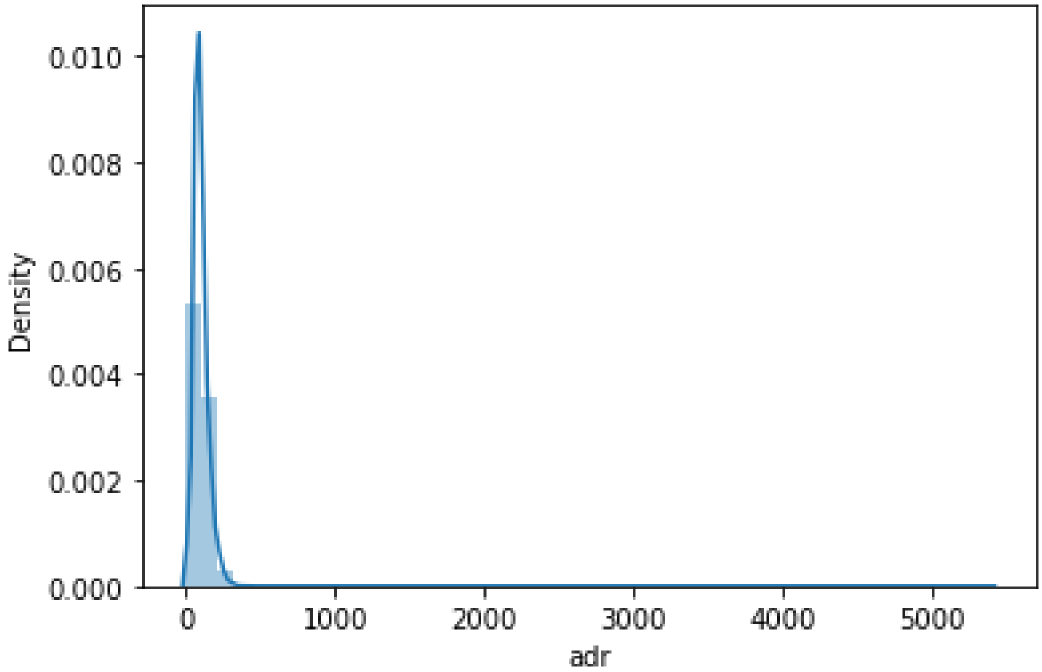
[157]:
```
sns.distplot(entire_df['lead_time'])
```

```
[157]: <AxesSubplot:xlabel='lead_time', ylabel='Density'>
```



```
[158]: sns.distplot(entire_df['adr'])
```

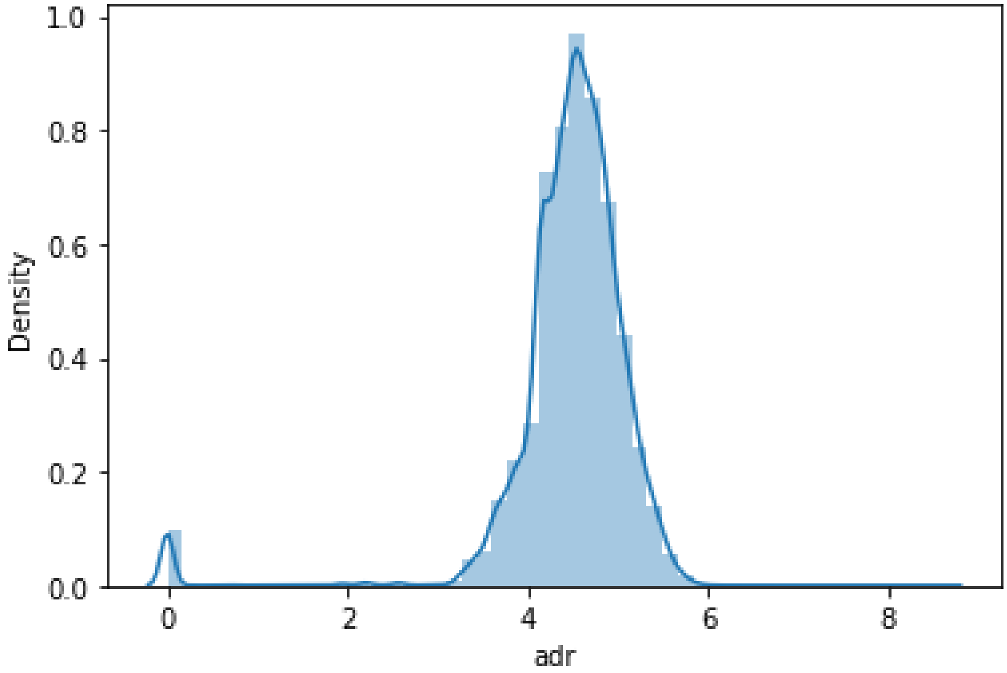
```
[158]: <AxesSubplot:xlabel='adr', ylabel='Density'>
```



```
[159]: handle_outlier('adr')
```

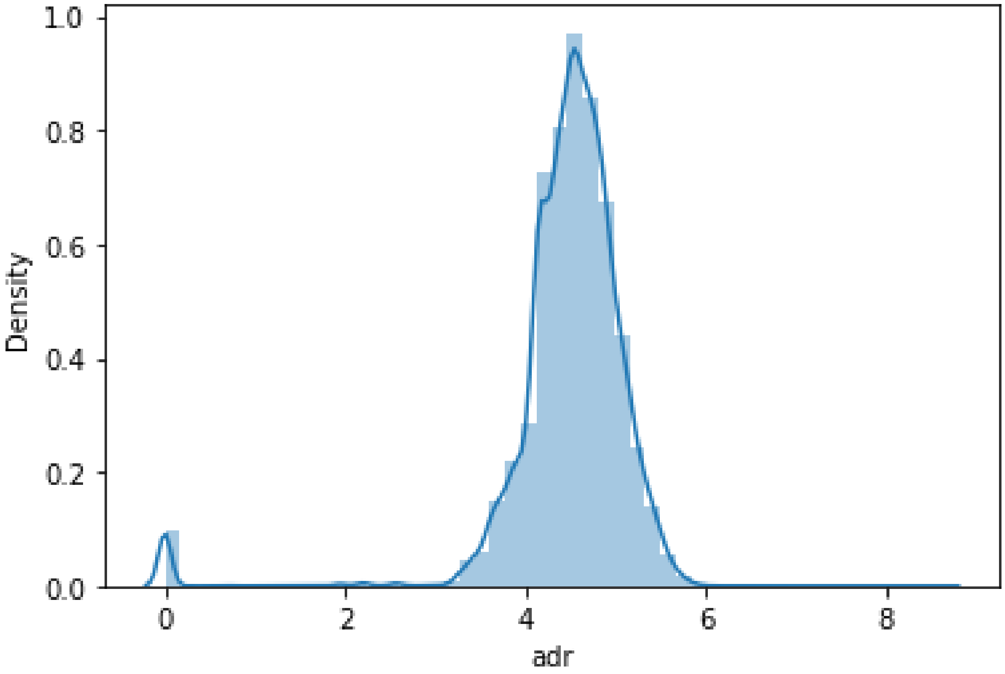
```
[160]: sns.distplot(entire_df['adr'].dropna())
```

```
[160]: <AxesSubplot:xlabel='adr', ylabel='Density'>
```



[161]:
```
sns.distplot(entire_df['adr'])
```

[161]:
```
<AxesSubplot:xlabel='adr', ylabel='Density'>
```



[162]:
```
#applying feature importance
#most important features

entire_df.isnull().sum()
```

[162]:
```
hotel                                0
arrival_date_month                   0
meal                                 0
market_segment                       0
distribution_channel                 0
reserved_room_type                   0
deposit_type                         0
customer_type                        0
year                                 0
month                                0
day                                  0
is_canceled                          0
lead_time                            0
```

```
arrival_date_year                 0
arrival_date_week_number          0
arrival_date_day_of_month         0
stays_in_weekend_nights           0
stays_in_week_nights              0
adults                            0
children                          0
babies                            0
is_repeated_guest                 0
previous_cancellations            0
previous_bookings_not_canceled    0
booking_changes                   0
agent                             0
company                           0
days_in_waiting_list              0
adr                               1
required_car_parking_spaces       0
total_of_special_requests         0
dtype: int64
```

[163]:
```
entire_df.dropna(inplace=True)
```

[164]:
```
#dependent and independent feature

#dependent feature
y = entire_df['is_canceled']

#independent features
x=entire_df.drop('is_canceled',axis=1)
```

[165]:
```
from sklearn.linear_model import Lasso
from sklearn.feature_selection import SelectFromModel
```

[166]:
```
feature_selmodel = SelectFromModel(Lasso(alpha=0.005, random_state=0))
```

[167]:
```
feature_selmodel.fit(x,y)
```

[167]:
```
SelectFromModel(estimator=Lasso(alpha=0.005, random_state=0))
```

[168]:
```
feature_selmodel.get_support()
```

[168]:
```
array([False, False, False, False, False, False,  True, False,  True,
        True,  True,  True,  True,  True,  True,  True,  True, False,
        True, False, False,  True,  True,  True, False,  True,  True,
        True,  True,  True])
```

[169]:
```
cols = x.columns
```

[170]:

```
selected_feat = cols[feature_selmodel.get_support()]
```

[171]:

```
print('total features {}'. format(x.shape[1]))
print('selected features {}'.format (len(selected_feat)))
```

```
total features 30
selected features 19
```

[172]:

```
x = x[selected_feat]
```

4 logistic regression

[173]:

```
#applying machine learning
#cross validation of data
```

[174]:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

[188]:

```
X_train, X_test, y_train, y_test = train_test_split(x,y,test_size=0.25,␣
 ↪random_state=42)
```

[189]:

```
logreg = LogisticRegression()
```

[190]:

```
logreg.fit(X_train,y_train)
```

[190]:

```
LogisticRegression()
```

[191]:

```
y_pred = logreg.predict(X_test)
```

[192]:

```
y_pred
```

[192]:

```
array([0, 0, 0, …, 0, 1, 0], dtype=int64)
```

[184]:

```
from sklearn.metrics import confusion_matrix
```

[185]:

```
confusion_matrix(y_test, y_pred)
```

[185]:

```
array([[15450,  3292],
       [ 5181,  5925]], dtype=int64)
```

[193]:

```
from sklearn.metrics import accuracy_score
```

[194]:

```
accuracy_score(y_test,y_pred)
```

[194]:

```
0.7235995711605467
```

[195]:

```
from sklearn.model_selection import cross_val_score
```

```
[196]: score = cross_val_score(logreg, x, y, cv=10)
```

```
[197]: score.mean()
```

```
[197]: 0.6832624512461718
```

5 applying various algorithmn on this data.

```
[199]: #importing the models from sklearn

from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
```

```
[200]: #initializing the model

models = []

models.append(('LogisticRegression', LogisticRegression()))
models.append(('Naive bayes', GaussianNB()))
models.append(('RandomForest', RandomForestClassifier()))
models.append(('Decision tree ',DecisionTreeClassifier()))
models.append(('KNN',KNeighborsClassifier()))
```

```
[201]: #fit the models

for name,model in models:
    print(name)
    model.fit(X_train, y_train)

    predictions = model.predict(X_test)

    from sklearn.metrics import confusion_matrix
    print(confusion_matrix(predictions, y_test))
    print('\n')

    print(accuracy_score(predictions, y_test))
    print('\n')
```

```
LogisticRegression
[[16661  6204]
 [ 2046  4937]]


0.7235995711605467
```

```
Naive bayes
[[ 7091   767]
 [11616 10374]]


0.5851313320825516


RandomForest
[[18647   997]
 [   60 10144]]


0.9645872420262664


Decision tree
[[18188   591]
 [  519 10550]]


0.9628115786652373


KNN
[[18640   901]
 [   67 10240]]


0.9675690163495042
```

```
[ ]:
```